

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Comugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 25, 2004, 04:58:59, Search time 4328 Seconds

(without alignments)
11036.050 Million cell updates/sec

Title: US-10-037-417-45

Perfect score: 1102

Sequence: 1 gggccctgctcctgggcat.....ggggttcgatgggcatcc 1102

Scoring table: IDENTITY NUC

Gapop 10.0, Gapext 1.0

Searched: 3470272 seqs, 21671516995 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : GenBml:*

1: gb_ba:*

2: gb_hg:*

3: gb_in:*

4: gb_om:*

5: gb_ov:*

6: gb_pat:*

7: gb_ph:*

8: gb_pl:*

9: gb_pr:*

10: gb_ro:*

11: gb_sts:*

12: gb_sy:*

13: gb_un:*

14: gb_vl:*

15: gb_da:*

16: em_fun:*

17: em_hum:*

18: em_in:*

19: em_mu:*

20: em_om:*

21: em_ov:*

22: em_ov:*

23: em_pat:*

24: em_ph:*

25: em_pl:*

26: em_ro:*

27: em_sts:*

28: em_un:*

29: em_vl:*

30: em_hg_hum:*

31: em_hg_inv:*

32: em_hg_other:*

33: em_hg_mus:*

34: em_hg_pla:*

35: em_hg_rod:*

36: em_hg_mam:*

37: em_hg_vrt:*

38: em_sy:*

39: em_hg_hum:*

40: em_hg_mus:*

41: em_hggo_other:*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Query Length	DB ID	Description
1	1102	100.0	1102	6	AX512287
2	1102	100.0	1102	6	AX512289
3	815.6	74.0	2457	6	AX360096
4	815.6	74.0	2681	6	AX480935
5	806.6	73.2	2810	6	BD127529
6	806.6	73.2	2810	9	AK075142
7	535.4	48.6	537	6	AX342934
8	472.8	42.9	670	6	BD125219
9	472.8	42.9	670	6	BD126485
10	442.8	40.2	456	6	AX342936
11	280.2	25.4	127769	9	AC009088
12	280.2	25.4	195476	2	AC135044
13	232.4	21.1	1796	6	AX098193
14	232.4	21.1	1800	9	BC001462
15	232.4	21.1	1835	6	AX098215
16	226.4	20.5	944	6	AX375744
17	225.6	20.5	1020	6	AX823252
18	225.6	20.5	1718	9	BC062334
19	225.6	20.5	1755	9	BC036846
20	224	20.3	1606	6	AX676264
21	224	20.3	1613	9	AR256990
22	224	20.3	1613	9	AF536382
23	221.2	20.1	1834	6	AX335777
24	221.2	20.1	1834	6	AX336076
25	221.2	20.1	1834	6	AX336340
26	221.2	20.1	1834	6	AX409567
27	221.2	20.1	1834	6	AX474697
28	221.2	20.1	1834	9	HUMPROS
29	220.2	20.0	71177	2	AC101465
30	220.2	20.0	185788	10	AC124461
31	220.2	20.0	219763	10	AC093175
32	217.8	19.8	232119	2	AC106629
33	217.8	19.8	244161	2	AC117170
34	217	19.7	64328	2	AC124816
35	216.8	19.7	1130	6	AR256996
36	205.2	18.6	786	6	AX375740
37	200	18.1	1727	10	AY335911
38	200	18.1	1743	10	BC003851
39	200	18.1	1856	10	AB038244
40	200	18.1	1870	10	AF378085
41	195.2	17.7	1797	10	AF188613
42	187.8	17.0	1360	10	AY262280
43	185.6	16.8	882	6	AX675583
44	183.2	16.6	1142	6	AR219285
45	183.2	16.6	1169	6	AR219284

ALIGNMENTS

RESULT 1

AX512287

LOCUS

DEFINITION

AX512287

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

Homo sapiens (human)

Homo sapiens

Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.

REFERENCE

1

Kekuda, R., Alsobrook, J. P., Tcherny, V. T., Liu, X., Sytek, K. A., Patrajan, M., Grose, W. M., Lepley, D. M., Burgess, C. E., Verne, C. A., Li, L., Gorman, L., Edinger, S., Scire, P., Ellerman, K., Malyankar, U.,

Rothenberg, M., Stone, D., Boldog, F., Shenoy, S., and Anderson, D.
 Proteins and nucleic acids encoding same
 Patent: WO 02053742-A 43 11-JUL-2002;
 Curagen Corporation (US)
 FEATURES
 source
 1. .1102
 /organism="Homo sapiens"
 /mol_type="unassigned DNA"
 /db_xref="taxon:9606"
 ORIGIN
 Query Match 100.0%; Score 1102; DB 6; Length 1102;
 Best Local Similarity 100.0%; Pred. No. 2.6e-164;
 Matches 1102; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 GGGGCCCTGCTCTGGGCCATGCGCAGAGAGGGGCTCTGGGGCCCTGGGGCAGCTGGGGGCT 60
 Db 1 GGGGCCCTGCTCTGGGCCATGCGCAGAGAGGGGCTCTGGGGCCCTGGGGCAGCTGGGGGCT 60
 QY 61 GTGGCCAAATCTGACTCATATCACTTTACGGGTTGGTGCCTCCGGACCCCGCTAGGGGC 120
 Db 61 GTGGCCAAATCTGACTCATATCACTTTACGGGTTGGTGCCTCCGGACCCCGCTAGGGGC 120
 QY 121 CCGCGTACTGCGGGCCCTGAGCCCTCGGCCGCGCATGCTGGGGGCTCAAACCGCGAG 180
 Db 121 CCGCGTACTGCGGGCCCTGAGCCCTCGGCCGCGCATGCTGGGGGCTCAAACCGCGAG 180
 QY 181 CCGGGCACTGCTGGGCAAGTGGAGCTGACCTGACCATGGAGTGGCCACATCTGCGGGGC 240
 Db 181 CCGGGCACTGCTGGGCAAGTGGAGCTGACCTGACCATGGAGTGGCCACATCTGCGGGGC 240
 QY 241 TCCCTCATGCCCCCTCTGCTGCTCTCTCCGTCTCTCTGCTCTCTCTCTCTCTCTCTCTCT 300
 Db 241 TCCCTCATGCCCCCTCTGCTGCTCTCTCTCCGTCTCTCTGCTCTCTCTCTCTCTCTCTCT 300
 QY 301 TTGGAGCCCGCGGCGAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 360
 Db 301 TTGGAGCCCGCGGCGAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 360
 QY 361 GACGGCGGCGCACACCGCGGCACTGCTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 420
 Db 361 GACGGCGGCGCACACCGCGGCACTGCTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 420
 QY 421 GAGCTGGGCGCGACCTGGCCCTGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCT 480
 Db 421 GAGCTGGGCGCGACCTGGCCCTGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCT 480
 QY 481 GTGTGGCTGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 540
 Db 481 GTGTGGCTGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 540
 QY 541 GCCACCGGCTGGGGAGAGCTCCAGAGGAGAGTCTCTGCTCTCTCTCTCTCTCTCTCTCTCT 600
 Db 541 GCCACCGGCTGGGGAGAGCTCCAGAGGAGAGTCTCTGCTCTCTCTCTCTCTCTCTCTCTCT 600
 QY 601 GAGTGGAGCTTAAGCTGCTGGCGAGGCGACCTGTCAATGTCTTACAGCAGCCCGGT 660
 Db 601 GAGTGGAGCTTAAGCTGCTGGCGAGGCGACCTGTCAATGTCTTACAGCAGCCCGGT 660
 QY 661 CCCTTCAACCTCACTCTCCAGATATGCCAGGATGCTGTGCTGTGCTGTGCTGTGCTGTGCT 720
 Db 661 CCCTTCAACCTCACTCTCCAGATATGCCAGGATGCTGTGCTGTGCTGTGCTGTGCTGTGCT 720
 QY 721 CGCAGGACACCTGCGAGGGTGACTCTGGGGGGCCCTGGTCTGTGAGGAGGCGGGCGCG 780
 Db 721 CGCAGGACACCTGCGAGGGTGACTCTGGGGGGCCCTGGTCTGTGAGGAGGCGGGCGCG 780
 QY 781 TGGTTCAGAGGAGGAATCAACAGCTTTGGGTTGGTGTGGACGAGAGAAACCGCCCTGGA 840
 Db 781 TGGTTCAGAGGAGGAATCAACAGCTTTGGGTTGGTGTGGACGAGAGAAACCGCCCTGGA 840
 QY 841 GTTTTCACTGCTGGCTACCTATCAGGATGATACGGAGCAGGTGATGGGTTTCAGAG 900
 Db 841 GTTTTCACTGCTGGCTACCTATCAGGATGATACGGAGCAGGTGATGGGTTTCAGAG 900

Db 841 GTTTTCACTGCTGGCTACCTATCAGGATGATACGGAGCAGGTGATGGGTTTCAGAG 900
 QY 901 COTGGGCTGCTTTTCCACCAGCCCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 960
 Db 901 COTGGGCTGCTTTTCCACCAGCCCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 960
 QY 961 GCATTCTCTGGATCTGCGCAGAACTTTTGGGCGCTTGTCCCATATATCAGTAGAGTTC 1020
 Db 961 GCATTCTCTGGATCTGCGCAGAACTTTTGGGCGCTTGTCCCATATATCAGTAGAGTTC 1020
 QY 1021 TCAACTGGGACCAAAAGCCTTGTCTCCCTCCCTGGCTCTCTCCACACTCTCTCTCTGGGCTC 1080
 Db 1021 TCAACTGGGACCAAAAGCCTTGTCTCCCTCCCTGGCTCTCTCCACACTCTCTCTCTGGGCTC 1080
 QY 1081 TGGGGGCTCTGATGGGGGCTCC 1102
 Db 1081 TGGGGGCTCTGATGGGGGCTCC 1102
 RESULT 2
 AX512289
 LOCUS AX512289 1102 bp DNA linear PAT 27-SEP-2002
 DEFINITION Sequence 45 from Patent WO02053742.
 ACCESSION AX512289
 VERSION AX512289.1 GI:23392632
 KEYWORDS
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE
 1
 AUTHORS
 Rothenberg, M., Stone, D., Boldog, F., Shenoy, S., and Anderson, D.
 TITLE
 JOURNAL
 Patent: WO 02053742-A 43 11-JUL-2002;
 Curagen Corporation (US)
 FEATURES
 source
 1. .1102
 /organism="Homo sapiens"
 /mol_type="unassigned DNA"
 /db_xref="taxon:9606"
 ORIGIN
 Query Match 100.0%; Score 1102; DB 6; Length 1102;
 Best Local Similarity 100.0%; Pred. No. 2.6e-164;
 Matches 1102; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 GGGGCCCTGCTCTGGGCCATGCGCAGAGAGGGGCTCTGGGGCCCTGGGGCAGCTGGGGGCT 60
 Db 1 GGGGCCCTGCTCTGGGCCATGCGCAGAGAGGGGCTCTGGGGCCCTGGGGCAGCTGGGGGCT 60
 QY 61 GTGGCCAAATCTGACTCATATCACTTTACGGGTTGGTGCCTCCGGACCCCGCTAGGGGC 120
 Db 61 GTGGCCAAATCTGACTCATATCACTTTACGGGTTGGTGCCTCCGGACCCCGCTAGGGGC 120
 QY 121 CCGCGTACTGCGGGCCCTGAGCCCTCGGCCGCGCATGCTGGGGGCTCAAACCGCGAG 180
 Db 121 CCGCGTACTGCGGGCCCTGAGCCCTCGGCCGCGCATGCTGGGGGCTCAAACCGCGAG 180
 QY 181 CCGGGCACTGCTGGGCAAGTGGAGCTGACCTGACCATGGAGTGGCCACATCTGCGGGGC 240
 Db 181 CCGGGCACTGCTGGGCAAGTGGAGCTGACCTGACCATGGAGTGGCCACATCTGCGGGGC 240
 QY 241 TCCCTCATGCCCCCTCTGCTGCTCTCTCCGTCTCTCTGCTCTCTCTCTCTCTCTCTCTCT 300
 Db 241 TCCCTCATGCCCCCTCTGCTGCTCTCTCTCCGTCTCTCTGCTCTCTCTCTCTCTCTCTCT 300
 QY 301 TTGGAGCCCGCGGCGAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 360
 Db 301 TTGGAGCCCGCGGCGAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 360

361 GAGCGCGGCACACCCGCGCAGTGGCGGCCCATGCTGGTCCGCGCCCAACTACAGCCAAAGTG 420
Db GAGCGCGGCACACACCCGCGCAGTGGCGGCCCATGCTGGTCCGCGCCCAACTACAGCCAAAGTG 420
421 GAGCTGGGCGCGCAGCTGGCCCTGCTGGCGCTGGCCCTCAACCGCCAGCCCTGGGCGCCCGCC 480
Db GAGCTGGGCGCGCAGCTGGCCCTGCTGGCGCTGGCCCTCAACCGCCAGCCCTGGGCGCCCGCC 480
481 GTGTGGCCCTGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 540
Db GTGTGGCCCTGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 540
481 GGCACCGGCTGGGAGACGTCACAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 600
Db GGCACCGGCTGGGAGACGTCACAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 600
601 GAAGTGGAGCTAAGGCTGCTGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 660
Db GAAGTGGAGCTAAGGCTGCTGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 660
661 CCCTTCAACCTCACTCTCCAGATATTGCCAGGAGTCTGTGTGCTGCTGCTGCTGCTGCTGCTG 720
Db CCCTTCAACCTCACTCTCCAGATATTGCCAGGAGTCTGTGTGCTGCTGCTGCTGCTGCTGCTG 720
721 CGCAGGAGACACCTGCCAGGAGTCTGCTGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 780
Db CGCAGGAGACACCTGCCAGGAGTCTGCTGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 780
781 TGGTTCCAGGAGGAGGAGTCTGCTGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 840
Db TGGTTCCAGGAGGAGGAGTCTGCTGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 840
841 GTTTTCACTGCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 900
Db GTTTTCACTGCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 900
901 CTGGGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 960
Db CTGGGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 960
961 GCATTCTCTGATCTGCGAGATCCCTTTGAGGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1020
Db GCATTCTCTGATCTGCGAGATCCCTTTGAGGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1020
1021 TCACTGGGACCAAAAGCCTTGTCTCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1080
Db TCACTGGGACCAAAAGCCTTGTCTCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1080
1081 TGGGGGTTCTGATGGGCGCTCC 1102
Db TGGGGGTTCTGATGGGCGCTCC 1102

RESULT 3
AX360096
LOCUS AX360096 2457 bp DNA linear PAT 13-FEB-2002
DEFINITION Sequence 52 from Patent WO0200860.
ACCESSION AX360096
VERSION AX360096.1 GI:18675722
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE 1
AUTHORS Plowman, G., Whyte, D., Sudarsanam, S., Manning, G., Caenepeel, S. and Charyczak, G.
TITLE Novel proteases
JOURNAL Patent: WO 0200860-A 52 03-JAN-2002;
Sugen, Inc. (US)
FEATURES Location/Qualifiers
1..2457
/organism="Homo sapiens"

/mol_type="unassigned DNA"
/db_xref="taxon:9606"
ORIGIN
Query Match 74.0%; Score 815.6; DB 6; Length 2457;
Best Local Similarity 99.5%; Pred. No. 3.1e-119;
Matches 818; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
QY 124 CGTACTGCGGGCGCCCTGAGCCCTCGGCCCATGCTGGGGGGGCTCAACCGCGAGCCG 183
Db 106 CTGGAGTCTGGGGCGCCCTGAGCCCTCGGCCCATGCTGGGGGGGCTCAACCGCGAGCCG 165
QY 184 GGCACCTGGCCCTTGGCAAGTGAAGCTGACCATGGAGGTGGCCACATCTGCGGGGCTCC 243
Db 166 GGCACCTGGCCCTTGGCAAGTGAAGCTGACCATGGAGGTGGCCACATCTGCGGGGCTCC 225
QY 244 CTCATGCGCCCTCTGGGTCTCTCGCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 303
Db 226 CTCATGCGCCCTCTGGGTCT 285
QY 304 GAGCCCGCGCGGAGTGGTTCGCTACTGCTGGGCTGCTGCGGCTGCTGCGGAGGCGGCGCTGGAC 363
Db 286 GAGCCCGCGCGGAGTGGTTCGCTACTGCTGGGCTGCTGCGGCTGCTGCGGAGGCGGCGCTGGAC 345
QY 364 GCGCGGCACACCCGCGCAGTGGCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 423
Db 346 GCGCGGCACACCCGCGCAGTGGCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 405
QY 424 CTGGGCGCGGAGTGGCCCTGCTGCGCTGCTGCGCTCAACCGCGAGCTGGGCGGCGGCGCTG 483
Db 406 CTGGGCGCGGAGTGGCCCTGCTGCGCTGCTGCGCTCAACCGCGAGCTGGGCGGCGGCGCTG 465
QY 484 TGGCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 543
Db 466 TGGCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 525
QY 544 ACCGCTGCGGAGAGCTCCAGAGGAGATCCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 603
Db 526 ACCGCTGCGGAGAGCTCCAGAGGAGATCCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 585
QY 604 GTGGAGCTAAGGCTGCTGGGCGAGGCGCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 663
Db 586 GTGGAGCTAAGGCTGCTGGGCGAGGCGCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 645
QY 664 TTCAACCTCACTCTCCAGATATTGCCAGGAGTGTGTGCTGCTGCTGCTGCTGCTGCTGCTG 723
Db 646 TTCAACCTCACTCTCCAGATATTGCCAGGAGTGTGTGCTGCTGCTGCTGCTGCTGCTGCTG 705
QY 724 AGGACACCTGCGAGGAGTGTGTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 783
Db 706 AGGACACCTGCGAGGAGTGTGTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 765
QY 784 TTCCAGGCGAGGAGTCAACAGGCTTGGGCTTGGGCTGCTGCGGAGGAGGAGGAGGAGGAGG 843
Db 766 TTCCAGGCGAGGAGTCAACAGGCTTGGGCTTGGGCTGCTGCGGAGGAGGAGGAGGAGGAGG 825
QY 844 TTCACTGCTGTGGCTACCTATGAGGAGTGGATACCGGAGGAGGAGTGTGCTGCTGCTGCTG 903
Db 826 TTCACTGCTGTGGCTACCTATGAGGAGTGGATACCGGAGGAGGAGTGTGCTGCTGCTGCTG 885
QY 904 GGGCGCTGCTTTCACCCAGCCCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 945
Db 886 GGGCGCTGCTTTCACCCAGCCCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 927
RESULT 4
AX480935
LOCUS AX480935 2681 bp DNA linear PAT 12-AUG-2002
DEFINITION Sequence 31 from Patent WO0246383.
ACCESSION AX480935
VERSION AX480935.1 GI:22217574
KEYWORDS Homo sapiens (human)
SOURCE

ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1
REFERENCE
AUTHORS Yue.H., Azimzai.Y., Kallick.D.A., Baughn.M.R., Griffin.J.A.,
Swarnakar.A., Lal.P.G., Walla.N.K., Hafalia.A.J., Gandhi.A.R.,
Au-Young.J., Elliott.V.S., Ramkumar.J., Thangavelu.K., Lu.Y.,
Warren.B.A., Lu.D.A., Lee.E.A., Tribouley.C.M., Arvizu.C.,
Delegeane.A.M., Yao.M.G., Khan.F.A. and Sanjanwala.M.M.
Protein modification and maintenance molecules
Patent: WO 0246383-A 31.13-JUN-2002;
Incyte Genomics, Inc. (US)
FEATURES
Location/Qualifiers
source
1..2681
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"
/note="Incyte ID No: 2751509CB1"
ORIGIN
Query Match 74.0%; Score 815.6; DB 6; Length 2681;
Best Local Similarity 99.5%; Pred. No. 3e-119;
Matches 818; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
QY 124 CGGTACTGCGGGCGCCCTGAGCCCTCGGCCGATCGTGGGGGCTCAAACGCGAGCGG 183
Db 106 CTGAGCTCGGGCGCCCTGAGCCCTCGGCCGATCGTGGGGGCTCAAACGCGAGCGG 165
QY 184 GGCACCTGGCCCTTGGCAAGTGCAGCCTGCACCATGAGGTGGCCACATCTGCGGGGCTCC 243
Db 166 GGCACCTGGCCCTTGGCAAGTGCAGCCTGCACCATGAGGTGGCCACATCTGCGGGGCTCC 225
QY 244 CTCATCGCCCTCTCTGGGTCTCTCCGCTGCTCACTGTTTTCATGCAATGGGCGTTG 303
Db 226 CTCATCGCCCTCTCTGGGTCTCTCCGCTGCTCACTGTTTTCATGCAATGGGCGCTG 285
QY 304 GAGCCGCGCGGCGAGTGGTCTGCTGCTGGCGTGCACCTCCAGGACGGGCGCTGGAC 363
Db 286 GAGCCGCGCGGCGAGTGGTCTGCTGCTGGCGTGCACCTCCAGGACGGGCGCTGGAC 345
QY 364 GCGCGGCACACCCGCGCAGTGGCGCGCATCTGTTGGTGGCGGCACTACAGCCAACTGGAG 423
Db 346 GCGCGGCACACCCGCGCAGTGGCGCGCATCTGTTGGTGGCGGCACTACAGCCAACTGGAG 405
QY 424 CTGGGCGGCGAGCTGGCCCTCTGCGCTGCGCTGCTGCGCTGCGCTGCGCTGCGCTG 483
Db 406 CTGGGCGGCGAGCTGGCCCTCTGCGCTGCGCTGCGCTGCGCTGCGCTGCGCTGCG 465
QY 484 TGGCCTGTGCTGCTGCGCGCGCTCACACCGCTTCTGTCAGCGCACCGCTGCGGCG 543
Db 466 TGGCCTGTGCTGCTGCGCGCGCTCACACCGCTTCTGTCAGCGCACCGCTGCGGCG 525
QY 544 ACCGCTGGGGAGAGCTCAGAGGAGATCTCTGCTCTCCCTCTCCCTGGTGTACAGGAA 603
Db 526 ACCGCTGGGGAGAGCTCAGAGGAGATCTCTGCTCTCCCTCTCCCTGGTGTACAGGAA 585
QY 604 GTGGAGCTAAGCTGCTGGCGAGGCCACCTGCTCAATGCTCTACAGCCAGCCGCTCC 663
Db 586 GTGGAGCTAAGCTGCTGGCGAGGCCACCTGCTCAATGCTCTACAGCCAGCCGCTCC 645
QY 664 TTCAACCTCACTCTCCAGATATTGCCAGGATGCTGTGTGCTGCTGCTGCTGCTGCT 723
Db 646 TTCAACCTCACTCTCCAGATATTGCCAGGATGCTGTGTGCTGCTGCTGCTGCTGCT 705
QY 724 AGGACACCTGCGCAGGTGACTCTGGGGGCGCCCTGGTGTGTGAGAAAGGCGCGCTGG 783
Db 706 AGGACACCTGCGCAGGTGACTCTGGGGGCGCCCTGGTGTGTGAGAAAGGCGCGCTGG 765
QY 784 TTCCAGGAGGAATCACAGCTTTGGTTTGGCTGTGGACGAGAAACCGCCCTCGAGTT 843
Db 766 TTCCAGGAGGAATCACAGCTTTGGCTTTGGCTGTGGACGAGAAACCGCCCTCGAGTT 825
844 TTCACTGTGTGGCTACCTATGAGGCATGGATACGGGAGAGGTGATGGGTTTCAGAGCCT 903

Db 826 TTCACTGTGTGGCTACCTATGAGGCATGGATACGGGAGCAGTGTGGTTTCAGAGCCT 885
QY 904 GGGCTGCTTTCCACCAGCCAGCCAGAGACCCAGTCAGAT 945
Db 886 GGGCTGCTTTCCACCAGCCAGCCAGAGACCCAGTCAGAT 927
RESULT 5
BD127529 2810 bp DNA linear PAT 18-SEP-2002
LOCUS Primer for synthesizing full-length cDNA and use thereof.
DEFINITION BD127529
ACCESSION BD127529
VERSION BD127529.1 GI:23224474
KEYWORDS JP 2002017375-A/2960.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 2810)
Ota.T., Nishikawa.T., Isogai.T., Hayashi.K., Ishii.S., Kawai.Y.,
Wakamatsu.A., Sugiyama.T., Negai.K., Kojima.S., Otsuki.T. and
Koga.H.
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
OS Homo sapiens (human)
PN JP 2002017375-A/2960
PD 22-JAN-2002
PF 07-JUL-2000 JP 200253172
PI TOSHIO OTA,TETSUO NISHIKAWA,TAKAO ISOGAI,KOJI HAYASHI,SHIZUKO
PI YURI KAWAI,AI WAKAMATSU,TOMOYASU SUGIYAMA,KEIICHI NAGAI, PI
SHINICHI KOJIMA,
PI TETSUJI OTSUKI,HISASHI KOGA
PC
C12N15/09,C07K14/47,C07K16/18,C12N1/15,C12N1/19,C12N1/21,C12N5/PC
10,
PC C12P21/02,C12Q1/68//C12P21/08,G06F17/30,C12N15/00,C12N5/00 CC
Primer for synthesizing full-length cDNA and use thereof FH Key
FT CDS Location/Qualifiers
300)..(2597).
source
1..2810
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
ORIGIN
Query Match 73.2%; Score 806.6; DB 6; Length 2810;
Best Local Similarity 92.0%; Pred. No. 7.9e-118;
Matches 862; Conservative 0; Mismatches 74; Indels 1; Gaps 1;
QY 10 TCCTGGGCATGGCCAGAGGGGCTCTGGGGCTGGGAGCTGGGGCTGTGGCCAT 69
Db 23 TCTAGCCATGGCCCGGACCTGCTCTCCCTTGTGATGCTGTATCAGTCCATC 82
QY 70 TCTGACTCATCTACTTTACGGGTGTGGTGGCTGGGAGCCGCTAGGGGCCCCCGTAC 129
Db 83 CCAGGAGCTTCCAGGACTCAGCTCTCAGTCTACCCAGGAGAACTGAAGATCTGGAC 142
QY 130 TGGGGGCGCCCTGAGCCCTTGGCCCGCATGCTTGGGGGCTCAAAACGCGAGCGGGGAC 188
Db 143 TGGGGGCGCCCTGAGCCCTTGGCCCGCATGCTGGGGGCTCAAAACGCGAGCGGGGAC 202
QY 189 CTGGCTTGGCACTGAGCCTGCACCATGGAGGTGGCCACATCTGCGGGGCTCCCTCAT 248
Db 203 CTGGCTTGGCACTGAGCCTGCACCATGGAGGTGGCCACATCTGCGGGGCTCCCTCAT 262
QY 249 GCGCCCTCTCTGGGTCTCTCCGCTGCTCACTGTTTTCATGCAATGGAGCTTGGAGCC 308
Db 263 GCGCCCTCTCTGGGTCTCTCCGCTGCTCACTGTTTTCATGCAATGGAGCTTGGAGCC 322


```
QY 309 CGCGCCGAGTGGTGGTACTGCTGGGCGTGCATCCACAGACGGGCCCTCTGACGGCGC 368
Db 323 CGCGCCGAGTGGTGGTACTGCTGGGCGTGCATCCACAGACGGGCCCTCTGACGGCGC 382
QY 369 GCACACCCGCGAGTGGCGCCATGCTGTGGCGGCACTACAGCAAGTGGAGCTGGG 428
Db 383 GCACACCCGCGAGTGGCGCCATGCTGTGGCGGCACTACAGCAAGTGGAGCTGGG 442
QY 429 CGCGACCTGGCCCTGCTGGGCGTGCATCCCGCCAGCCTGGGCGCCCGCGTGTGGCC 488
Db 443 CGCGACCTGGCCCTGCTGGGCGTGCATCCCGCCAGCCTGGGCGCCCGCGTGTGGCC 502
QY 489 TGCTGTGCTGGCCCGGCTCTACACCGCTTGTGTGACAGGACCGCTGTGTGGCGACCGG 548
Db 503 TGCTGTGCTGGCCCGGCTCTACACCGCTTGTGTGACAGGACCGCTGTGTGGCGACCGG 562
QY 549 CTGGGAGACGCTCCAGAGGAGATCTCTGCTCTCTCCCTGGGTGCTACAGAGTGA 608
Db 563 CTGGGAGACGCTCCAGAGGAGATCTCTGCTCTCTCCCTGGGTGCTACAGAGTGA 622
QY 609 GCTAAGGCTGTGGGCGAGCCACCTGTCAATGTCTCTACAGCCAGCCCGTCCCTTCAA 668
Db 623 GCTAAGGCTGTGGGCGAGCCACCTGTCAATGTCTCTACAGCCAGCCCGTCCCTTCAA 682
QY 669 CTTCACTCTCCAGATATTCACAGGATGCTGTGTGCTGCTACAGAGGCGCGAGGGA 728
Db 683 CTTCACTCTCCAGATATTCACAGGATGCTGTGTGCTGCTACAGAGGCGCGAGGGA 742
QY 729 CACTGTGCCAGGTGACTCTGGGGGGCCCTGTCTGTGAGGAGCGCGCTGTGCTTCA 788
Db 743 CACTGTGCCAGGTGACTCTGGGGGGCCCTGTGTGAGGAGCGCGCTGTGCTTCA 802
QY 789 GGCAGGAATCACCAGCTTTGGTGTGGTGTGGAGAGAAACCGCCCTGGAGTTTTCAC 848
Db 803 GGCAGGAATCACCAGCTTTGGTGTGGTGTGGAGAGAAACCGCCCTGGAGTTTTCAC 862
QY 849 TGCTGTGCTACTATGAGGATGATAGGAGGAGAGTGTGAGGTTACAGCCCTGGGCG 908
Db 863 TGCTGTGCTACTATGAGGATGATAGGAGGAGAGTGTGAGGTTACAGCCCTGGGCG 922
QY 909 TGCCCTTCCACCCAGCCCGCAGAACAGCCAGTCAGAT 945
Db 923 TGCCCTTCCACCCAGCCCGCAGAACAGCCAGTCAGAT 959

RESULT 6
AK075142
LOCUS
DEFINITION
Homo sapiens cDNA FLJ90661 fis, clone PLACE1005003, weakly similar
to PROSTASIN PRECURSOR (EC 3.4.21.-).
ACCESSION
AK075142
VERSION
AK075142.1 GI:22761040
KEYWORDS
oligo capping; fis (full insert sequence).
SOURCE
Homo sapiens (human)
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1
Isogai, T., Ota, T., Nishikawa, T., Hayashi, K., Otsuki, T.,
Sugiyama, T., Suzuki, Y., Nagai, K., Sugano, S., Ishii, S.,
Kawai-Hio, Y., Saito, K., Yanamoto, J., Wakamatsu, A., Nakamura, Y.,
Kojima, S., Nagahari, K., Masuho, Y., Ono, T., Okano, K., Yoshikawa, Y.,
Aotsuka, S., Sasaki, N., Hattori, A., Okumura, K., Iwayanagi, T. and
Ninomiya, K.
NEDO human cDNA sequencing project
Unpublished
2 (bases 1 to 2810)
Isogai, T. and Otsuki, T.
Direct Submission
Submitted (25-MAR-2002) Takao Isogai, Helix Research Institute,
Genomics Laboratory; 1532-3 Yana, Kisarazu, Chiba 292-0812, Japan
(E-mail: genomics@hri.co.jp, Tel: 81-438-52-3975, Fax: 81-438-52-3986)
NEDO human cDNA sequencing project supported by Ministry of
COMMENT
```

Economy, Trade and Industry of Japan; cDNA full insert sequencing: Research Association for Biotechnology; cDNA library construction: Institute of Medical Science, University of Tokyo, Laboratory of Genome Structure, Human Genome Center; cDNA 5' - & 3' - end one pass sequencing and clone selection: Helix Research Institute (supported by Japan Key Technology Center etc.).

FEATURES

source

```
1..2810
/organism="Homo sapiens"
/mo_type="mRNA"
/db_xref="taxon:9606"
/clone="PLACE1005003"
/tissue_type="placenta"
/clone_lib="PLACE1"
/note="cloning vector: pME18SFL3"
300..2600
/note="unnamed protein product"
/codon_start=1
/protein_id="BAC11431.1"
/db_xref="GI:22761041"
```

CDS

```
1
translation="MTNCTLEPAEWSVLLGVHSDGPDLDGAHTFAVAIVPANYSQ
VELGADLALLSLASPASLGPVWPVCLPRASHRFVHGTACNATMGWDVQEDPLPLPW
VLQVELLLGATCCQCHYSGPFPNLTQLILPGMLCAGYEGRRDITCGSGGGLVC
EGGWFOAGITSTFGCGRRNRPFVTAATVATWIREQVWGSEPFAPFTQPKQK
SDQPRENCTIALPECGKAPRPAQWVAQVMPGSRPCHGALVSESWLAPASCF
LDPNSSPPRDLDAWRVLLPSRAERVARLVHONASWDNASDLALLQLRTPNLS
AASRPVCLPHEHYFLPGSRCLARWGEALGEGALLAEALLGGMWCHLYGEGA
AVPLQTHGPWLSHVTRGAYLEDQLADWGPDEETETCTCPPTPHSHGACGURLEAPV
GVLMFWLAHVHAGDVCTGILLAPWVAATHCVLRFGSTTPVTVYVLRAGASSL
PQGHQRLVISIRLPHQLPPLALLESLRVPESALPCLHPHAPLGGASCWV
LGWKEPDORVPAVAASILTQRIICDLVQGLIPGTLVLAEGDENCEMTSAPLL
COMTEGWSLLVGMVAQVSRRELFAALPREEAMISQTVGEANLLPPSGSPHWPVGGSNLC
PPELAKAGSPPHAYVFLLLLLTLLQS"
```

ORIGIN

```
Query Match 73.2%; Score 806.6; DB 9; Length 2810;
Best Local Similarity 92.0%; Pred. No. 7.9e-118;
Matches 862; Conservative 0; Mismatches 74; Indels 1; Gaps 1;

QY 10 TCCTGGGCGCATGGCCAGAGGGGTCTCTGGGGCTGGGCGAGCTGGGGGCTGTGGCCAAAT 69
Db 23 TCTTAGCGCATGGGCGCGCACCTGCTCTCCCTCTTGTGATGCTTGTCTCATCAGTCCCATC 82
QY 70 TCTGACTCATACTCACTTTACGGGTGTGCTGCGGTCCGACCCGCTAGGGGCGCCCGGTAC 129
Db 83 CCAGGAGCTTCCAGGACTCAGCTCTAGTCTCTACCCAGGAGAACCTGAAGATCTGAC 142
QY 130 TGGGCGCGCTGAGCCCTCGGCGCGCATGCT-GGGGGGCTCAAAACGCGAGCGCGGCGAC 188
Db 143 TSCGGGCGCCCTGAGCCCTCGGCGCGCATCTGCGGGGGGCTCAAAACGCGAGCGCGGCGAC 202
QY 189 CTGCGCTTGGCAAGTACGCTGACCACTGAGGTGGCCACATCTCGGGGGCTCCCTCAT 248
Db 203 CTGCGCTTGGCAAGTACGCTGACCACTGAGGTGGCCACATCTGCGGGGGCTCCCTCAT 262
QY 249 CGCCCTCTCTCGGTCTCTCCGCTGCTCACTGTTTTCATGACGAATGGAGCTGTGGAGCC 308
Db 263 CGCCCTCTCTCGGTCTCTCTCCGCTGCTCACTGTTTTCATGACGAATGGAGCTGTGGAGCC 322
QY 309 CGCGCGCGAGTGTGCTGTGCTGTGGGGGTGCACTCCCGAGACGGGCGCCCTGTGGAGCGG 368
Db 323 CGCGCGCGAGTGTGCTGTGCTGTGGGGGTGCACTCCCGAGACGGGCGCCCTGTGGAGCGG 382
QY 369 GCACACCGCGCAGTGGCGCGCATCTGCTGGTGGCGCCCACTACAGCCCAAGTGGAGCTGGG 428
Db 383 GCACACCGCGCAGTGGCGCGCATCTGCTGGTGGCGCCCACTACAGCCCAAGTGGAGCTGGG 442
QY 429 CGCGGACCTGGCCCTGTGCTGGGCTTGGGCTTACCCCGCGCAGCTGGGCGCCCGCGTGTGGCC 488
Db 443 CGCGGACCTGGCCCTGTGCTGGGCTTGGGCTTACCCCGCGCAGCTGGGCGCCCGCGTGTGGCC 502
QY 489 TGTCTGCTGCCCCCGCGCTCACACCGCTTGTGTGCAAGGACCGCCCTGTGTGGGCGACCGG 548
```

Db 503 TGTCTGCTGCTCCCGCGCTCACACCGCTTCTGTCAGCGCACCCGCTGCTGGCCACCGG 562
QY 549 CTGGGGAGACCTCCAGGAGGAGATCTCTGCTCTCCCTCTGCTGCTACAGGAAGTGA 608
Db 563 CTGGGGAGACCTCCAGGAGGAGATCTCTGCTCTCCCTCTGCTGCTACAGGAAGTGA 622
QY 609 GCTAAGGCTGCTGGGCGAGGCCACTCTCAATGCTCTACAGCCAGCCGCTGCTCAA 668
Db 623 GCTAAGGCTGCTGGGCGAGGCCACTCTCAATGCTCTACAGCCAGCCGCTGCTCAA 682
QY 669 CCTCACTCTCAGATATTCAGGAGGATGCTGTGCTGCTACCGAGGCGCGAGGA 728
Db 683 CCTCACTCTCAGATATTCAGGAGGATGCTGTGCTGCTACCGAGGCGCGAGGA 742
QY 729 CACCTGCCAGGGTGACTCTTGGGGGGCCCTGCTGTGAGGAAGCGCGCTGTTCCA 788
Db 743 CACCTGCCAGGGTGACTCTTGGGGGGCCCTGCTGTGAGGAAGCGCGCTGTTCCA 802
QY 789 GGCAGGAATACACAGCTTTGGGTTGGTGCTGTGAGCGAGAAACCGCCCTGAGTTTCA 848
Db 803 GGCAGGAATACACAGCTTTGGGTTGGTGCTGTGAGCGAGAAACCGCCCTGAGTTTCA 862
QY 849 TGCTGTGCTACCTATGAGGATGATACGAGGAGCTGATGGTTTCAGAGCTGGGCC 908
Db 863 TGCTGTGCTACCTATGAGGATGATACGAGGAGCTGATGGTTTCAGAGCTGGGCC 922
QY 909 TGCTTTTCCACCCAGCCCGCCAGAGACCCAGTCAGAT 945
Db 923 TGCTTTTCCACCCAGCCCGCCAGAGACCCAGTCAGAT 959
RESULT 7
AX342934/c
LOCUS AX342934 537 bp DNA linear PAT 12-JAN-2002
DEFINITION Sequence 1 from Patent WO0198467.
ACCESSION AX342934
VERSION AX342934.1 GI:18152213
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1
Xiao, Y. and Morozov, V.
Regulation of human prostatic-like serine protease
Patent: WO 0198467-A 1 27-DEC-2001;
Bayer Aktiengesellschaft (DE)
JOURNAL
FEATURES
Location/Qualifiers
source 1..537
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"
ORIGIN
Query Match 48.6%; Score 535.4; DB 6; Length 537;
Best Local Similarity 99.8%; Pred. No. 7e-75;
Matches 536; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 243 CCTCATCGCCCTCCCTGGGTCCTCTCGGCTGCTCACTGTTTCATGACGAATGGAGCTT 302
Db 537 CCTCATCGCCCTCCCTGGGTCCTCTCGGCTGCTCACTGTTTCATGACGAATGGAGCTT 478
QY 303 GGAGCCGCGCGCCGAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 362
Db 477 GGAGCCGCGCGCCGAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 418
QY 363 GGGCGCGACACCGCGAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 422
Db 417 GGGCGCGACACCGCGAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 358
QY 423 GCTGGGCGCGACCTGGGCTGCTGGGCTGCTGGGCTGCTGGGCTGCTGGGCTGCTGGGCTG 482

Db 357 GCTGGGCGCGACCTGGGCTGCTGGGCTGCTGGGCTGCTGGGCTGCTGGGCTGCTGGGCT 298
QY 483 GTGGCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 542
Db 297 GTGGCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 238
QY 543 CACCGGCTGGGGAGAGCTCCAGGAGGAGATCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 602
Db 237 CACCGGCTGGGGAGAGCTCCAGGAGGAGATCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 178
QY 603 AGTGGAGCTAAGCTGCTGGGCGAGGCGACCTGCTCAATGCTCTACAGCCAGCCGCTGCTGCT 662
Db 177 AGTGGAGCTAAGCTGCTGGGCGAGGCGACCTGCTCAATGCTCTACAGCCAGCCGCTGCTGCT 118
QY 663 CTTCAACCTCACTCTCCAGATATTCAGGAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 722
Db 117 CTTCAACCTCACTCTCCAGATATTCAGGAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 58
QY 723 CAGGAGACCTGCTCCAGGCTGACTCTGGGGGGCCCTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 779
Db 57 CAGGAGACCTGCTCCAGGCTGACTCTGGGGGGCCCTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 1
RESULT 8
BD125219
LOCUS BD125219 670 bp DNA linear PAT 18-SEP-2002
DEFINITION Primer for synthesizing full-length cDNA and use thereof.
ACCESSION BD125219
VERSION BD125219.1 GI:23220164
KEYWORDS JP 2002017375-A/650.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 670)
Oca, T., Nishikawa, T., Isogai, T., Hayashi, K., Ishii, S., Kawai, Y.,
Wakamatsu, A., Sugiyama, T., Nagai, K., Kojima, S., Otsuki, T. and
Koga, H.
Primer for synthesizing full-length cDNA and use thereof
Patent: JP 2002017375-A 650 22-JAN-2002;
JOURNAL HELIX RESEARCH INSTITUTE
COMMENT OS Homo sapiens (human)
PN JP 2002017375-A/650
PD 22-JAN-2002
PF 07-JUL-2000 JP 200253172
PI TOSHIO OTA, TETSUO NISHIKAWA, TAKAO ISOGAI, KOJI HAYASHI, SHIZUKO
PI ISHII,
PI YURI KAWAI, AI WAKAMATSU, TOMOYASU SUGIYAMA, KEIICHI NAGAI, PI
SHINICHI KOJIMA,
PI TETSUJI OTSUKI, HISASHI KOGA
PC C12N15/09, C07K14/47, C07K16/18, C12N1/15, C12N1/19, C12N1/21, C12N5/10,
PC C12P21/02, C12Q1/68//C12P21/08, G06F17/30, C12N15/00, C12N5/00 CC
Primer for synthesizing full-length cDNA and use thereof FH Key
Location/Qualifiers
FT source 1..670
FT /organism="Homo sapiens (human)"
FEATURES
Location/Qualifiers
source 1..670
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
ORIGIN
Query Match 42.9%; Score 472.8; DB 6; Length 670;
Best Local Similarity 86.9%; Pred. No. 4.9e-65;
Matches 563; Conservative 0; Mismatches 81; Indels 4; Gaps 4;
QY 10 TCCTGGGCCCATGGCCCGAGAGGGGGTCTGGGGCTCTGGGCGAGCTGGGGGCTGGGCAAT 69
Db 23 TCCTAGGCCCATGGCCCGAGAGCTGCTCTCCCTTGTGATGCTTCTCATCAGTCCCATC 82

HQLDSYSEDKAVSTLKDIIIPHSYLOQSGODIALLOLSRPITPSYRPICTLPAANA
SFPNGLHCTVTGWGHVAPSVLLPKPQLOLEVLPIKRETCNCLNYIDAKPEBPHVQ
EDMVCAGVGGKADCGDGGPLSCDVEGLWLTGIVSWGDACGNRPVYTLAS
YASHIQKVTELQPRVVPQESQSDNLCGSHLAFSPAQAQGLLRPLFLPIGLALG
LLSPWLSEH
305..1018
/note="Tryp SPC; Region: Trypsin-like serine protease"
/db_xref="CDD:smart0020"

misc_feature

ORIGIN

Query Match 21.1%; Score 232.4; DB 9; Length 1800;
Best Local Similarity 55.0%; Pred. No. 3.1e-27;
Matches 508; Conservative 0; Mismatches 401; Indels 15; Gaps 2;

QY 1 GGGCCCTGTCTGGCCCATCGCCAGAGGGGCTCTGGGGCTGGGCGCTGGGGCT 60
DB |||||
QY 158 GGGCCCTGTCTGGCCCATCGCCAGAGGGGCTCTGGGGCTGGGCGCTGGGGCT 217
DB |||||
QY 61 GTGGCCAAATTCGACTCATCTCACTTTACG3GTTGGTGCCTGTCGGAACCGCTAGGGGC 120
DB |||||
QY 218 GTGGCCATTCGCTCTATCTTGGATTACTCCGGTCGGGACAGAGCGGAGGGGCAGAA 277
DB |||||
QY 121 CCCCGTACTCGGGGCGCCCTGAGCCCTCGGCCCGCATCTGGGGGGCTCAAAACGCGAG 180
DB |||||
QY 278 GCTCCCTGCGGTGTG-----GCCCCCAAGCAGCATCACAGGTGGCAGAGTGCAGTC 331
DB |||||
QY 181 CGGGACCTGGCCCTGGCAAGTGAAGCTGACCATGAGAGTGGCCACATCTGGGGGGGC 240
DB |||||
QY 332 GCGGTGAGTGGCCCTGGCAGGTGAGCATCACTATGAAAGGGTCCATGTGTGTGGTGGC 391
DB |||||
QY 241 TCCCTCAFGCCGCCCTCTGGTCTCTCCGCTGCTACTGTTTCATGACGAAATGGGAGC 300
DB |||||
QY 392 TCTCTCGTGTCTGAGCAGTGGTGTGTGCTGCTCACTGCTCTTCCCGCAGCGA----- 444
DB |||||
QY 301 TTGAGCCCGGGCGAGTGGTGTGCTGCTGCTGGCGTGGCTATCCAGGACGGGCCCTG 360
DB |||||
QY 445 --GCACCAAAAGAACCTATGAGGTCAAGTGGGGGCCCAACAGCTAGCTCTCTACTTCC 502
DB |||||
QY 361 GACGGCGGCACACCCGGCAGTGGCCGCCATCGTGGTGGCGGCCAATACAGGCCAAGTG 420
DB |||||
QY 503 GAGNAGCCAGAGTCAACCTGAGGACATCATCCCCACCCAGCTACTCTCAGAGAG 562
DB |||||
QY 421 GAGTGGCGCGCAGCTGGCCCTGTGGCCCTGAGCTCAACCGGCAGCTGGGGCCCGCC 480
DB |||||
QY 563 GGTCCCGAGGCGCATTTGCACTCTCTCAACTCAGCAGACCCATCACCTTCTCCCGCTAC 622
DB |||||
QY 481 GTGTGGCTGTCTGCTGCGCCCGCGCTCACACCTTGTGTGCGAGCGACCGCTGTGG 540
DB |||||
QY 623 ATCCGGCCCATCTGCTCCCTGACGCCAAGCGCTCTTCCCAACGGGCTCCACTGCACT 682
DB |||||
QY 541 GCCACCGGTGGGAGACGTCCAGGAGCGAGATCCTCTGCTCTCTCCCTGGGTGCTACAG 600
DB |||||
QY 683 GTCACTGTGCTGGGTCAATGTGGCCCTCAGTAGAGCTCTGACGCCCTGACGCCCTGAG 742
DB |||||
QY 601 GAAGTGGAGTAAAGCTGTGGGCGAGGCCACTGTCAATGTCTTACAGCAGCGCCCGGT 660
DB |||||
QY 743 CAACCTCGAGGTGCTCTGATCACTGCTGAGAGCTGTAACTGCTGCTGTACAAATCACGCC 802
DB |||||
QY 661 CCCTTCAACTCACTCTCCAGATATTCAGGAGTGTGTGTGGTGTGCTGTGCTGTGAGG 720
DB |||||
QY 803 AAGCTGAGGAGCCGACCTTGTCCAAAGGACATGGTGTGTGCTGTGCTGTGAGGGG 862
DB |||||
QY 721 CCGAGGACACCTGCGAGGCTGACTCTGGGGGGCCCTGGTGTGTGAGGAGCGCGCCGCG 780
DB |||||
QY 863 GGCAGGAGCGCTGCGAGGCTGACTCTGGGGGCCCACTCTCTGCTGCTGTGGAGGTCTC 922
DB |||||
QY 781 TGGTTCAGGAGGAATACAGCTTTGGTGTGGCTGTGGAGCGAGAAACCGCCCTGGA 840
DB |||||
QY 923 TGGTACCTGACGGGCATTGTGAGCTGGGAGATGCTGTGGGGGCCCGCAACAGGCGCTGT 982
DB |||||
QY 841 GTTTTCACTGTGTGGTCACTTATGAGGATGATACGGGAGCAGGTGATGGTTTCAGAG 900
DB |||||
QY 983 GTGTACACTCTGCGCTCCAGCTATGCTCTCTGGATCCAAAGCAAGGTGACAACTCCAG 1042
DB |||||

QY 901 CCTGGGCTGCTTTTCCACCCAG 924
DB 1043 CCTGCTGTGGTCCCAAAACCCAG 1066

RESULT 15

AX098215 1835 bp DNA linear PAT 30-MAR-2001

LOCUS AX098215

DEFINITION Sequence 127 from Patent WO0118542.

ACCESSION AX098215

VERSION AX098215.1 GI:13515326

KEYWORDS

SOURCE Homo sapiens (human)

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE

1 Lee, J., Thompson, P. and Lillie, J.

Identification, assessment, prevention, and therapy of ovarian

cancer

Patent: WO 0118542-A 127 15-MAR-2001;

Millennium Predictive Medicine, Inc. (US)

JOURNAL

Location/Qualifiers

1. 1835

/organism="Homo sapiens"

/mol_type="unassigned DNA"

/db_xref="taxon:9606"

FEATURES

source

ORIGIN

Query Match 21.1%; Score 232.4; DB 6; Length 1835;
Best Local Similarity 55.0%; Pred. No. 3.1e-27;
Matches 508; Conservative 0; Mismatches 401; Indels 15; Gaps 2;

QY 1 GGGCCCTGTCTGGCCCATCGCCAGAGGGGCTCTGGGGCTGGGCGCTGGGGCT 60
DB |||||
QY 212 GGGCCCTGTCTGGCCCATCGCCAGAGGGGCTCTGGGGCTGGGCGCTGGGGCT 271
DB |||||
QY 61 GTGGCCAAATTCGACTCATCTCACTTTACGGGTTGGTGCCTGCGGACCGCGTAGGGGC 120
DB |||||
QY 272 GTGGCCATTCGCTCTATCTTGGATTACTCGGTGCGGGACAGAGCGGAGGGGCAGAA 331
DB |||||
QY 121 CCCCGTACTCGGGGCGCCCTGAGCCCTCGGCCCGCATCTGGGGGGGTCAAAACGCGAG 180
DB |||||
QY 332 GCTCCCTGCGGTGTG-----GCCCCCAAGCAGCATCACAGGTGGCAGAGTGCAGTC 385
DB |||||
QY 181 CCGGCGACCTGGCTTGGCAAGTCACTGCACTGCACTGAGGTGGCCACATCTGGGGGGC 240
DB |||||
QY 386 GCGGTCAGTGGCCCTGGAGGTGAGCATCACTATGAGGCGTCCATGTGTGTGGTGGC 445
DB |||||
QY 241 TCCCTCAATCGCCCTCTCTGGGTCTCTCTCGCTGCTCACTGTTTCAAGCAATGGGAGC 300
DB |||||
QY 446 TCTCTGCTGTCTGAGCAGTGGTGTCTGCTCACTGCTCTCTCTCTCTCTCTCTCTCT 498
DB |||||
QY 301 TTGAGCCCGCGCGGAGTGTGTGGTACTGTGGGCTGCACTCCAGGACGGGCCCTG 360
DB |||||
QY 499 --GCACCAAAAGAACCTATGAGGTCAAGTGGGGGCCCAACAGCTAGACTCTCTACTTCC 556
DB |||||
QY 361 GACGGCGGCACACCGCGAGTGGCGCCATCTGTGTGGTGGCGGCCAATACAGCCAAAGTG 420
DB |||||
QY 557 GAGNAGCCAGAGTCAAGCCTGAAGGACATATCCCCACCCAGCTACTCTCAGAGAG 616
DB |||||
QY 421 GAGTGGCGCGCACTGGCCCTCTGCGCTGCGCTTACCCGCGAGCCTGGGGCCCGCC 480
DB |||||
QY 617 GGTCCCAAGGGCGCATTTGCACTCTTCCAACTCAGCAGACCCATCACCTTCTCCCGTAC 676
DB |||||
QY 481 GTGTGGCTGTCTGCTGCCCGCGCTCAACCGCTTGTGCGAGCGCACCGCCCTGTGG 540
DB |||||
QY 677 ATCCGGCCCATCTGCTCTCTCTGCAAGCCCAACCGCTCTTCCCAACCGGCTCCACTGCA 736
DB |||||
QY 541 GCCACCGGCTGGGAGAGCTCCAGGAGCAGATCCTCTGCTCTCTCTCTCTCTCTCTCTCT 600
DB |||||
QY 737 GTCACTGCTGGGTGTCATGTGGCCCCCTCAGTGAGCTCTCTGACGCCCAAGCCACTGAG 796
DB |||||

QY	601	GAAGTGGAGCTAAGGCTGCTGGGCGAGGCCACCTGTCAATGHTCTACAGCCAGCCCGGT	660
Db	797	CAACTCGAGGTGCTCTGTATCAGTCGTGAGACGTGTAACTGCCTGTACACATCGACGCC	856
QY	661	CCCTTCAACCTCACTCTCCAGATATTGCCAGGGATGCTGTGTGCTGTGCTACCCAGAGGGC	720
Db	857	AAGCCTGAGGAGCCGACATTGTCCAGAGGACATGCTGTGTGCTGTGCTATGTGAGGGG	916
QY	721	CGCAGGGACACCTGCCAGGGTGACTCTGGGGGCCCTGTGTGTGTGAGGAAGGCCGCCG	780
Db	917	GGCAGAGACCCCTGCCAGGGTGACTCTGGGGGCCCACTCTCCCTGTGTGAGGGTCTC	976
QY	781	TGGTTCAGSCAGGAATCACAGCTTTGGGTTTGGCTGTGGACGGAGAAACCGCCCTGGA	840
Db	977	TGGTACCTGACGGGCATTGTGAGCTGGGGAGATGCCCTGTGGGGCCCGCAACAGGCCCTGT	1036
QY	841	GTTTTCACTGCTGTGGTACCTATGAGGCATGATACGGGAGCAGGTGATGGGTTCAAG	900
Db	1037	GTGTACACTGTGGCCTCCAGCTATGCTCTGGATCCAAAGCAAGGTGACAGAACTCCAG	1096
QY	901	CCTGGGCTGCCCTTCCACCCAG	924
Db	1097	CTCTGTGTGTGCCCCAAACCCAG	1120

Search completed: February 25, 2004, 10:31:04
Job time : 4339 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 24, 2004, 23:37:42 ; Search time 474 Seconds
(without alignments)
9876.612 Million cell updates/sec

Title: US-10-037-417-45

Perfect score: 1102

Sequence: 1 gggccctgtctggccat.....ggggtctgatggggccccc 1102

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 3373863 seqs, 2124099041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N_Geneseq_29Jan04:*

1: geneseqn1980s:*

2: geneseqn1990s:*

3: geneseqn2000s:*

4: geneseqn2001as:*

5: geneseqn2001bs:*

6: geneseqn2002s:*

7: geneseqn2003as:*

8: geneseqn2003bs:*

9: geneseqn2003cs:*

10: geneseqn2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	1102	100.0	1102	6	ABQ93901 Human pro
2	1102	100.0	1102	6	ABQ93902 Human pro
3	815.6	74.0	2457	6	ABK31795 DNA encod
4	815.6	74.0	2681	6	ABQ75956 Human PMM
5	806.6	73.2	2810	4	Aak94500 Human ful
6	535.4	48.6	537	6	ABK13565 Human pro
7	472.8	42.9	670	4	Aak92190 Human CDN
8	472.8	42.9	670	4	Aak93456 Human pro
9	442.8	40.2	456	6	ABK13566 Human pro
10	433.4	39.3	597	5	Aas69040 DNA encod
11	232.4	21.1	1736	5	Aaf98698 Human ova
12	232.4	21.1	1835	5	Aaf98720 Human lat
13	232.4	21.1	3382	6	ABZ35336 Human gen
14	230.8	20.9	1733	7	ABT31936 Human bre
15	226.4	20.5	944	6	ABA94396 Human pro
16	225.6	20.5	843	6	ABQ86175 Novel hum
17	225.6	20.5	849	6	ABQ86176 Novel hum
18	224	20.3	1606	6	ABs67730 DNA encod
19	224	20.3	1613	4	AAS01360 Human ser
20	221.2	20.1	1834	6	ABs76501 cDNA enco
21	221.2	20.1	1834	6	ABL67949 Ovary can
22	221.2	20.1	1834	6	ABL68512 Kidney ca
23	221.2	20.1	1834	6	ABL68248 Kidney ca

24	221.2	20.1	1834	6	ABK12241	ABK12241 cDNA enco
25	221.2	20.1	1834	6	ABN95716	ABn95716 Gene #221
26	220.6	20.0	1668	3	AAC77814	Human can
27	216.8	19.7	1130	4	AAS01366	Recombina
28	205.2	18.6	786	6	ABA94394	Human pro
29	205.2	18.6	2122	5	ABA94394	Human pro
30	195.2	17.7	1797	7	ABs56430	Mouse cha
31	185.6	16.8	882	6	ABN85394	Partial H
32	184	16.7	882	6	ABN85395	Partial H
33	183.2	16.6	1142	3	AAC87796	Activatio
34	183.2	16.6	1142	4	AAC87796	Activatio
35	183.2	16.6	1169	3	AAC87795	Activatio
36	183.2	16.6	1169	4	AAC87795	Activatio
37	180.8	16.4	768	4	AAC87795	Activatio
38	179	16.2	1161	6	ABN85393	Human NOV
39	176.2	16.0	1327	2	AAC34034	Human EST
40	176.2	16.0	1327	3	AAC34034	Human EST
41	176.2	16.0	1327	3	AAC34034	Human EST
42	176.2	16.0	1327	7	ACD42567	Novel hum
43	176.2	16.0	1327	7	ACD42567	Novel hum
44	176.2	16.0	1327	7	ACA71766	Human sec
45	176.2	16.0	1327	7	ABX92406	Human PRO

ALIGNMENTS

RESULT 1

ABQ93901

ID ABQ93901 standard; DNA; 1102 BP.

XX

AC ABQ93901;

XX

DT 01-NOV-2002 (first entry)

XX

DE Human prostatin precursor-like NOV14a DNA, SEQ ID NO:43.

XX

Human; NOVA; neurological disorder; Alzheimer's disease;
Huntington's disease; Parkinson's disease; pain; behavioural disorder;
addiction; tuberculous sclerosis; cancer; immune disorder; allergy;
autoimmune disease; myasthenia gravis; aschma; arthritis; diabetes;
thyroiditis; cardiovascular disease; hypertension; reproductive disorder;
endometriosis; incontinence; psoriasis; scleroderma; alopecia; ulcer;
pancreatitis; cirrhosis; glomerular endochelosis; bacterial infection;
polycystic kidney disease; endocrine disorder; obesity; cardiomyopathy;
atherosclerosis; cell signal processing-related disorder;
metabolic pathway regulation disorder; cytostatic; neuroprotective;
antiinflammatory; immunosuppressive; analgesic; antiatherosclerotic;
dermatological; antibacterial; antiarthritic; hepatotropic; neurogenesis;
differentiation; proliferation; motility; haematopoiesis; wound healing;
angiogenesis; forensic biology; transgenic animal; drug screening;
gene therapy; NOV14a; prostatin precursor-like; chromosome 16; gene; ds.

XX

OS Homo sapiens.

XX

Key Location/Qualifiers

5'UTR 1..18

CDS /*tag= a

19..1050

/*tag= b

/partial

/product= "NOV14a"

/note= "No stop codon given in the specification"

XX

WO200253742-A2.

XX

PD 11-JUL-2002.

XX

PF 07-JAN-2002; 2002WO-US000375.

XX

PR 05-JAN-2001; 2001US-0260018P.

PR

PR 08-JAN-2001; 2001US-0260360P.

PR

PR 28-FEB-2001; 2001US-0272411P.

PR 02-MAR-2001; 2001US-0272817P.
PR 05-JUL-2001; 2001US-0303231P.
PR 12-JUL-2001; 2001US-0305060P.
PR 10-SEP-2001; 2001US-0318405P.
PR 12-SEP-2001; 2001US-0318700P.
PR 04-JAN-2002; 2002US-00037417.
XX
PA (CURA-) CURAGEN CORP.
XX
XX
PI Kekuda R, Alsobrook JP, Tchernev VT, Liu X, Spytek KA;
PI Patturajan M, Grosse WM, Lepley DM, Burgess CE, Vernet CAM, Li L;
PI Gorman L, Edinger S, Sciore P, Ellerman K, Malyankar U;
PI Rothenberg M, Stone D, Boldog F, Guo X, Shenoy S, Anderson D;
PI Padigaru M, Taupier RJ, Miller CE, Eisen A;
XX
DR WPI; 2002-583619/62.
DR P-PSDB; ABH09523.
XX
XX Novel polypeptides and nucleic acids homologous to transmembrane
PT receptor, thymosin, neuromodulin-like family of proteins for diagnosing,
PT treating cancer, atherosclerosis, neurological, skin and autoimmune
PT disorders.
XX
XX Claim 9a; Page 142; 323pp; English.
XX
XX The invention relates to 24 novel human proteins designated NOV1-NOV14
XX (ABB09501-ABB09524), collectively referred to as NOVX proteins, and
XX nucleic acids encoding them (ABQ93879-ABQ93902). NOVX proteins and
XX nucleotides are useful in the treatment, diagnosis or prevention of NOVX-
XX associated disorders or in the manufacture of a medicament for treating
XX such disorders, with specific applications described for each of the 24
XX NOVX proteins, based on their homology to known proteins. Various
XX disorders are associated with NOVX proteins including neurological
XX disorders (e.g., Alzheimer's, Huntington's and Parkinson's diseases),
XX pain, behavioural disorders, addiction, tuberosus sclerosis, cancers
XX (e.g., colorectal cancer, leukaemia and osteosarcoma), immune disorders
XX (e.g., allergies and autoimmune diseases), myasthenia gravis, asthma,
XX various forms of arthritis, diabetes, thyroiditis, cardiovascular disease
XX (e.g., hypertension), reproductive disorders, endometriosis,
XX incontinence, psoriasis, scleroderma, alopecia, ulcers, pancreatitis,
XX cirrhosis, glomerular endotheliosis, polycystic kidney disease, endocrine
XX disorders, obesity, bacterial infections and particularly cardiomyopathy,
XX atherosclerosis, cell signal processing-related disorders and disorders
XX of metabolic pathway regulation. NOVX nucleic acids and polypeptides may
XX be used to identify cellular receptors or downstream effectors which
XX binds to a NOVX protein, and are also useful as targets for the
XX identification of small molecules that modulate or inhibit processes such
XX as neurogenesis, cell differentiation, cell motility, cellular
XX proliferation, haematopoiesis, wound healing and angiogenesis. NOVX
XX nucleic acid sequences can be used to identify a cell or tissue type and
XX are useful as a source of primers or probes for forensic biology and for
XX identifying and cloning NOVX homologues in other cell types. Cells
XX comprising NOVX nucleic acids are useful for producing non-human
XX transgenic animals which are useful for studying the function and
XX activity of NOVX proteins and for identifying and evaluating modulators
XX of NOVX activity. The present sequence represents DNA encoding the
XX prostatic precursor-like protein NOV14a. The gene encoding NOV14a is
XX located on chromosome 16
SQ Sequence 1102 BP; 162 A; 371 C; 353 G; 216 T; 0 U; 0 Other;

Query Match 100.0%; Score 1102; DB 6; Length 1102;
Best Local Similarity 100.0%; Pred. No. 8.2e-185;
Matches 1102; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GGGCCCTTGTCCCTGGGCGCATGGCCAGAGGGGTCTCTGGGGCTGGGCGAGCTGGGGGCT 60
Db 1 GGGCCCTTGTCCCTGGGCGCATGGCCAGAGGGGTCTCTGGGGCTGGGCGAGCTGGGGGCT 60

QY 61 GTGGCCAAATTTCTGACTCATCTACTCACTTACGGGTGTGGTGGCGCTCGGACCCCGCTAGGGGC 120
Db 61 GTGGCCAAATTTCTGACTCATCTACTCACTTACGGGTGTGGTGGCGCTCGGACCCCGCTAGGGGC 120

RESULT 2
ABQ93902
ID ABQ93902 standard; DNA; 1102 BP.
XX

QY 121 CCCCCGTAATGGGGGCGCCCTGAGCCCTCGGCCCGCATCTGTGGGGGGCTCAACCGGCGAG 180
Db CCCCCGTAATGGGGGCGCCCTGAGCCCTCGGCCCGCATCTGTGGGGGGCTCAACCGGCGAG 180

QY 181 CCGGGCACCTGGGCTTGGCAAGTGAGCCTGACCATGAGAGGTGGCCACATCTCTCGGGGGC 240
Db CCGGGCACCTGGGCTTGGCAAGTGAGCCTGACCATGAGAGGTGGCCACATCTCTCGGGGGC 240

QY 241 TCCTCATCGCCCCCTCTCTGGGTCTCTCGGTGCTCACTGTCTTATGAGCAATGGGACG 300
Db TCCTCATCGCCCCCTCTCTGGGTCTCTCGGTGCTCACTGTCTTATGAGCAATGGGACG 300

QY 301 TTGGAGCCCGCGCCAGTGGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 360
Db TTGGAGCCCGCGCCAGTGGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 360

QY 361 GAGCGCGCGCACACCCGCGCAGTGGCGCCCATCTGTGTGCGCGCCCACTACTACAGCCAA 420
Db GAGCGCGCGCACACCCGCGCAGTGGCGCCCATCTGTGTGCGCGCCCACTACTACAGCCAA 420

QY 421 GAGCTGGCGCCCGACCTTGGCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 480
Db GAGCTGGCGCCCGACCTTGGCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 480

QY 481 GTGTGGCTCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 540
Db GTGTGGCTCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 540

QY 541 GCCACCGGCTGGGGAGACGCTCCAGGAGGAGATCTCTGCTGCTCTCTCTCTCTCTCTCT 600
Db GCCACCGGCTGGGGAGACGCTCCAGGAGGAGATCTCTGCTGCTCTCTCTCTCTCTCTCT 600

QY 601 GAAGTGGAGCTAAGCTGTGGGGGAGCGCCACCTGTCAATGTCTCTACAGCCAGCCCGGT 660
Db GAAGTGGAGCTAAGCTGTGGGGGAGCGCCACCTGTCAATGTCTCTACAGCCAGCCCGGT 660

QY 661 CCCTTCAACCTCACTCTCCAGATATTCGCGAGGATGCTGTGCTGCTGCTGCTGCTGCTG 720
Db CCCTTCAACCTCACTCTCCAGATATTCGCGAGGATGCTGTGCTGCTGCTGCTGCTGCTG 720

QY 721 CGCAGGACACCTGCCAGGGTGAATCTCTGGGGGGGGCCCTGTGCTGTGAGGAAGGGGGCG 780
Db CGCAGGACACCTGCCAGGGTGAATCTCTGGGGGGGGCCCTGTGCTGTGAGGAAGGGGGCG 780

QY 781 TGGTTCAGGAGGAAATCACAGCTTTGGGTTTGGCTGTGAGCGGAGAAACCGCCCTGGA 840
Db TGGTTCAGGAGGAAATCACAGCTTTGGGTTTGGCTGTGAGCGGAGAAACCGCCCTGGA 840

QY 841 GTTTTCACTGTGGCTTACCTATGAGGCATGGATACGGGAGCAGGTGATGGGTTTCAGAG 900
Db GTTTTCACTGTGGCTTACCTATGAGGCATGGATACGGGAGCAGGTGATGGGTTTCAGAG 900

QY 901 CTTGGGCTGCTCTTCCACCCAGCCAGAACCCAGTCCAGATGTTGTTTACATCAAAACG 960
Db CTTGGGCTGCTCTTCCACCCAGCCAGAACCCAGTCCAGATGTTGTTTACATCAAAACG 960

QY 961 GCATTCTGGATCTGCGCAGATCCCTTTTGGGCGCTTGTCCCATATATCAGTAGAGTC 1020
Db GCATTCTGGATCTGCGCAGATCCCTTTTGGGCGCTTGTCCCATATATCAGTAGAGTC 1020

QY 1021 TCAACTGGGAGCCAAAGCCTTGTCTCTCCCTGGCTCTCTCCACACTCTCTCTCTGGGCGCTC 1080
Db TCAACTGGGAGCCAAAGCCTTGTCTCTCCCTGGCTCTCTCCACACTCTCTCTCTGGGCGCTC 1080

QY 1081 TGGGGGTTCTGATGGGGCTCC 1102
Db 1081 TGGGGGTTCTGATGGGGCTCC 1102

AC ABQ93902;
 XX 01-NOV-2002 (first entry)
 XX Human prostatic precursor-like NOV14b DNA, SEQ ID NO:45.
 DE
 XX Human; NOVX; neurological disorder; Alzheimer's disease;
 KW Huntington's disease; Parkinson's disease; pain; behavioural disorder;
 KW addiction; tuberosus sclerosis; cancer; immune disorder; allergy;
 KW autoimmune disease; myasthenia gravis; asthma; arthritis; diabetes;
 KW thyroditis; cardiovascular disease; hypertension; reproductive disorder;
 KW endometriosis; incontinence; psoriasis; scleroderma; alopecia; ulcer;
 KW pancreatitis; cirrhosis; glomerular endotheliosis; bacterial infection;
 KW polycystic kidney disease; endocrine disorder; obesity; cardiomyopathy;
 KW atherosclerosis; cell signal processing-related disorder;
 KW metabolic pathway regulation disorder; cytostatic; neuroprotective;
 KW antiinflammatory; immunosuppressive; analgesic; antithrombotic;
 KW dermatological; antibacterial; antiarthritic; hepatotropic; neurogenesis;
 KW differentiation; proliferation; motility; haematopoiesis; wound healing;
 KW angiogenesis; forensic biology; transgenic animal; drug screening;
 KW gene therapy; NOV14b; prostatic precursor-like; gene; ds.
 XX
 OS Homo sapiens.
 XX
 XX Key Location/Qualifiers
 XX 5'UTR 1..18 /*tag= a
 FT CDS 19..1092 /*tag= b
 FT 3'UTR 1093..1102 /product= "NOV14b"
 FT /*tag= c
 XX
 XX W0200253742-A2.
 XX 11-JUL-2002.
 XX
 XX 07-JAN-2002; 2002WO-US000375.
 XX
 XX 05-JAN-2001; 2001US-0260018P.
 XX 08-JAN-2001; 2001US-0260360P.
 XX 28-FEB-2001; 2001US-0272411P.
 XX 02-MAR-2001; 2001US-0272817P.
 XX 05-JUL-2001; 2001US-0303231P.
 XX 12-JUL-2001; 2001US-0305060P.
 XX 10-SEP-2001; 2001US-0318405P.
 XX 12-SEP-2001; 2001US-0318700P.
 XX 04-JAN-2002; 2002US-00037417.
 XX
 XX (CURA-) CURAGEN CORP.
 XX
 XX Kekuda R, Alsobrook JP, Tchernev VT, Liu X, Spytek KA;
 PI Patturajan M, Grosse WM, Lepley DM, Burgess CE, Vernet CM, Li L;
 PI Gorman L, Edinger S, Sciore P, Ellerman K, Malyankar U;
 PI Rothenberg M, Stone D, Boldog F, Guo X, Shenoy S, Anderson D;
 PI Padigaru M, Taupier RJ, Miller CE, Eisen A;
 XX
 XX WPI: 2002-583619/62.
 XX P-PSDB; ABB09524.
 XX
 XX Novel polypeptides and nucleic acids homologous to transmembrane
 PT receptor, thymosin, neuromodulin-like family of proteins for diagnosing,
 PT treating cancer, atherosclerosis, neurological, skin and autoimmune
 PT disorders.
 XX
 XX Claim 9a; Page 143; 323pp; English.
 PS
 XX The invention relates to 24 novel human proteins designated NOV1-NOV14
 CC (ABB09501-ABB09524), collectively referred to as NOVX proteins, and
 CC nucleic acids encoding them (ABB093879-ABB093902). NOVX proteins and
 CC nucleotides are useful in the treatment, diagnosis or prevention of NOVX-
 CC associated disorders or in the manufacture of a medicament for treating
 CC such disorders, with specific applications described for each of the 24

CC NOVX proteins, based on their homology to known proteins. Various
 CC disorders are associated with NOVX proteins including neurological
 CC disorders (e.g., Alzheimer's, Huntington's and Parkinson's diseases),
 CC pain, behavioural disorders, addiction, tuberosus sclerosis, cancers
 CC (e.g., colorectal cancer, leukaemia and osteosarcoma), immune disorders
 CC (e.g., allergies and autoimmune diseases), myasthenia gravis, asthma,
 CC various forms of arthritis, diabetes, thyroditis, cardiovascular disease
 CC (e.g., hypertension), reproductive disorders, endometriosis,
 CC incontinence, psoriasis, scleroderma, alopecia, ulcers, pancreatitis,
 CC cirrhosis, glomerular endotheliosis, polycystic kidney disease, endocrine
 CC disorders, obesity, bacterial infections and particularly cardiomyopathy,
 CC atherosclerosis, cell signal processing-related disorders and disorders
 CC of metabolic pathway regulation. NOVX nucleic acids and polypeptides may
 CC be used to identify cellular receptors or downstream effectors which
 CC binds to a NOVX protein, and are also useful as targets for the
 CC identification of small molecules that modulate or inhibit processes such
 CC as neurogenesis, cell differentiation, cell motility, cellular
 CC proliferation, haematopoiesis, wound healing and angiogenesis. NOVX
 CC nucleic acid sequences can be used to identify a cell or tissue type and
 CC are useful as a source of primers or probes for forensic biology and for
 CC identifying and cloning NOVX homologues in other cell types. Cells
 CC comprising NOVX nucleic acids are useful for producing non-human
 CC transgenic animals which are useful for studying the function and
 CC activity of NOVX proteins and for identifying and evaluating modulators
 CC of NOVX activity. The present sequence represents DNA encoding the
 CC prostatic precursor-like protein NOV14b
 XX
 XX Sequence 1102 BP; 162 A; 371 C; 353 G; 216 T; 0 U; 0 Other;
 SQ
 Query Match 100.0%; Score 1102; DB 6; Length 1102;
 Best Local Similarity 100.0%; Pred. No. 8.2e-185;
 Matches 1102; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 GGGCCCTTGTCTCGGCGCATGGCCAGAGGGGGTCTCTGGGGCTGGGACGTGGGGCT 60
 DB 1 GGGCCCTTGTCTCGGCGCATGGCCAGAGGGGGTCTCTGGGGCTGGGACGTGGGGCT 60
 QY 61 GTGGCCCAATTTCTGACTCATACTACTTTTACGGGTGGTGGCGTCCGGACCCGCTAGGGGC 120
 DB 61 GTGGCCCAATTTCTGACTCATACTACTTTTACGGGTGGTGGCGTCCGGACCCGCTAGGGGC 120
 QY 121 CCCCGTACTCGGGGGCCCTGAGCCCTCGGCCCGCATCTGTGGGGGCTCAAAACGCGCAG 180
 DB 121 CCCCGTACTCGGGGGCCCTGAGCCCTCGGCCCGCATCTGTGGGGGCTCAAAACGCGCAG 180
 QY 181 CCGGCGACCTGGCCCTTGGCAAGTGGAGCTGCACCATGAGGTGGGCACATCTGCGGGGGC 240
 DB 181 CCGGCGACCTGGCCCTTGGCAAGTGGAGCTGCACCATGAGGTGGGCACATCTGCGGGGGC 240
 QY 241 TCCCTCATCGCCCCCTCTCTGGGTCTCTTCGCTGCTCACTGTTTCATGCAATGGGACG 300
 DB 241 TCCCTCATCGCCCCCTCTCTGGGTCTCTTCGCTGCTCACTGTTTCATGCAATGGGACG 300
 QY 301 TTGGAGCCCGCGCGAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 360
 DB 301 TTGGAGCCCGCGCGAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 360
 QY 361 GACGGCGGCACACCCCGCGCAGTGGCCGCCCATCTGTGTGTCGGGGCACTACAGCCAAAGTG 420
 DB 361 GACGGCGGCACACCCCGCGCAGTGGCCGCCCATCTGTGTGTCGGGGCACTACAGCCAAAGTG 420
 QY 421 GAGTGGCGCGGAGCTGGCCCTCTGTGGCGCTGGCCCTCACCCGCGAGCTGGGCCCGCCG 480
 DB 421 GAGTGGCGCGGAGCTGGCCCTCTGTGGCGCTGGCCCTCACCCGCGAGCTGGGCCCGCCG 480
 QY 481 GTGTGGCCCTGTCTGGCTGCGCCGCGCCTCACACCGCTTCGTGCGAGCGCCGCTGCTGG 540
 DB 481 GTGTGGCCCTGTCTGGCTGCGCCGCGCCTCACACCGCTTCGTGCGAGCGCCGCTGCTGG 540
 QY 541 GCACCCGCTGGGAGAGCTCCAGGAGGAGATCTCTGCTCTCTCCCTGGGTGTACAG 600
 DB 541 GCACCCGCTGGGAGAGCTCCAGGAGGAGATCTCTGCTCTCTCCCTGGGTGTACAG 600

QY 601 GAAGTGGAGCTAAGCTGTGGGGAGGCCACCTGTCAATGTCCTCTACAGCCAGCCCGGT 660
 Db 601 GAAGTGGAGCTAAGCTGTGGGGAGGCCACCTGTCAATGTCCTCTACAGCCAGCCCGGT 660
 QY 661 CCCTTCAACCTCACTCTCCAGATATTCAGGGAGTGTGTGTGCTACCCAGAGGGC 720
 Db 661 CCCTTCAACCTCACTCTCCAGATATTCAGGGAGTGTGTGTGCTACCCAGAGGGC 720
 QY 721 CGCAGGACACCTCCAGGGTGAATCTGGGGGGCCCTGGTCTGTGAGGAAGGGCGCGC 780
 Db 721 CGCAGGACACCTCCAGGGTGAATCTGGGGGGCCCTGGTCTGTGAGGAAGGGCGCGC 780
 QY 781 TGGTTCACAGGAGGAATCACCAGCTTTTGGCTGTGGACGGAGAAACCGCCCTGGA 840
 Db 781 TGGTTCACAGGAGGAATCACCAGCTTTTGGCTGTGGACGGAGAAACCGCCCTGGA 840
 QY 841 GTTTCACCTGTGTGCTACCTATGAGCATGGATACGGAGCAGGTGATGGTTCAGAG 900
 Db 841 GTTTCACCTGTGTGCTACCTATGAGCATGGATACGGAGCAGGTGATGGTTCAGAG 900
 QY 901 CCTGGGCTGCTTTCCACAGCCAGCCAGAGACCCAGTCAGATTGTTTACATCAAACG 960
 Db 901 CCTGGGCTGCTTTCCACAGCCAGCCAGAGACCCAGTCAGATTGTTTACATCAAACG 960
 QY 961 GCATTTCCTGGATCTGCAGATCCTTTTGGAGCCCTTGTCCCATATATAGTAGAGTC 1020
 Db 961 GCATTTCCTGGATCTGCAGATCCTTTTGGAGCCCTTGTCCCATATATAGTAGAGTC 1020
 QY 1021 TCAACTGGGACCAAAAGCTTGTCTCCCTGGCTCTCTCCACACTCTCTCTGGGCTC 1080
 Db 1021 TCAACTGGGACCAAAAGCTTGTCTCCCTGGCTCTCTCCACACTCTCTCTGGGCTC 1080
 QY 1081 TGGGGGTTCTGATGGGGCTCC 1102
 Db 1081 TGGGGGTTCTGATGGGGCTCC 1102

RESULT 3
 ID ABK31795
 XX ABK31795 standard; DNA; 2457 BP.
 AC ABK31795;
 XX
 DT 23-APR-2002 (first entry)
 XX
 DE DNA encoding novel human protease #52.
 XX
 KW Human; protease; cancer; immune-related disorder; cardiovascular disease;
 KW neuronal-associated disease; metabolic disorder; inflammatory disorder;
 KW nervous system disorder; sexual dysfunction; pain; mood disorder;
 KW hypertension; psychotic disorder; neurological disorder; dyskinesia;
 KW viral infection; human immunodeficiency virus; HIV; non-viral infection;
 KW ocular disease; cytostatic; gene; ds.
 XX
 OS Homo sapiens.
 XX
 PN WO200200860-A2.
 XX
 PD 03-JAN-2002.
 XX
 PF 26-JUN-2001; 2001WO-US020171.
 XX
 PR 26-JUN-2000; 2000US-0214047P.
 XX
 XX (SUGEN-) SUGEN INC.
 XX
 XX Plowman G, Whyte D, Sudarsanam S, Manning G, Caenepeel S;
 PI Charyczak G;
 XX
 XX WPI; 2002-139913/18.
 DR P-PSDB; AAU82753.
 XX
 XX Nucleic acids encoding novel human proteases, useful for

PT treating diseases and disorders such as cancers, immune-related diseases
 PT and disorders, cardiovascular disease (e.g. restenosis) and inflammatory
 PT disorders.
 XX Claim 26; Fig 1SS-TT; 313pp; English.
 XX
 XX The present invention relates to the isolation of novel human proteases,
 CC and the nucleic acids encoding them. The sequences of the invention are
 CC useful for treating diseases and disorders such as cancers (e.g. breast,
 CC colon, lung), immune-related diseases and disorders (e.g. inflammatory
 CC diseases and asthma), cardiovascular diseases (e.g. restenosis and
 CC coronary thrombosis), brain or neuronal-associated diseases, metabolic
 CC disorders (e.g. diabetes, obesity), inflammatory disorders (e.g.
 CC rheumatoid arthritis and psoriasis), central or peripheral nervous system
 CC diseases, migraines, pain, sexual dysfunction, mood disorders, attention
 CC disorders, neurological disorders, hypotension, hypertension, psychotic
 CC disorders, cognitive disorders (e.g. Alzheimer's disease, Parkinson's
 CC disease) and dyskinesias. The nucleic acids and polypeptides are also
 CC useful for treating viral infections caused by human immunodeficiency
 CC virus (HIV), and non-viral infections such as ocular disease (e.g.
 CC glaucoma) and macular degeneration. ABK31744-ABK31802 represent DNA
 CC sequences encoding for the novel human proteases of the invention
 XX
 SQ Sequence 2457 BP; 377 A; 851 C; 797 G; 432 T; 0 U; 0 Other;

Query Match 74.0%; Score 815.6; DB 6; Length 2457;
 Best Local Similarity 99.5%; Pred. No. 1.6e-134;
 Matches 818; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
 QY 124 CCGTACTCGGGCGCCCTGAGCCCTCGCGCCGATCGTGGGGGGCTCAAAACGCGCAGCGC 183
 Db 106 CTGAGCTGCGGAGCGCCCTGAGCCCTCGCGCCGATCGTGGGGGGCTCAAAACGCGCAGCGC 165
 QY 184 GGCACCTCTGGCCTTGGCAAGTGAGCTGCACCATGGAGTGGCCACATCTGCGGGGGTCC 243
 Db 166 GGCACCTGGCCTTGGCAAGTGAGCTGCACCATGGAGTGGCCACATCTGCGGGGGTCC 225
 QY 244 CTGATCGCCCTCTCTGGGTCTCTCGCTGCTCCTCTGTTTCATGACGAATGGGACGTG 303
 Db 226 CTGATCGCCCTCTCTGGGTCTCTCGCTGCTCCTCTGTTTCATGACGAATGGGACGTG 285
 QY 304 GAGCCGCGCGCGAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 363
 Db 286 GAGCCGCGCGCGAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 345
 QY 364 GCGCGCACACCCGCGCAGTGGCCCATCTGTTGGCGGCAACTACAGCAAGTGAG 423
 Db 346 GCGCGCACACCCGCGCAGTGGCCCATCTGTTGGCGGCAACTACAGCAAGTGAG 405
 QY 424 CTGGCGCGCGACCTGGCCCTCTGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 483
 Db 406 CTGGCGCGCGACCTGGCCCTCTGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 465
 QY 484 TGGCTGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 543
 Db 466 TGGCTGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 525
 QY 544 ACCGGCTGGGAGAGTCCAGAGCAGATCCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 603
 Db 526 ACCGGCTGGGAGAGTCCAGAGCAGATCCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 585
 QY 604 GTGGAGCTAAGGCTGCTGGGAGGCGCCACCTGTCAATGTCCTCTACAGCCAGCCCGTCC 663
 Db 586 GTGGAGCTAAGGCTGCTGGGAGGCGCCACCTGTCAATGTCCTCTACAGCCAGCCCGTCC 645
 QY 664 TTCAACCTCACTCTCCAGATATTCAGGGAGTGTGTGTGCTGCTGCTGCTGCTGCTGCTG 723
 Db 646 TTCAACCTCACTCTCCAGATATTCAGGGAGTGTGTGTGCTGCTGCTGCTGCTGCTGCTG 705
 QY 724 AGGGACACCTGCGAGGGTGAATCTGGGGGGCCCTGCTGCTGCTGCTGCTGCTGCTGCTG 783
 Db 706 AGGGACACCTGCGAGGGTGAATCTGGGGGGCCCTGCTGCTGCTGCTGCTGCTGCTGCTG 765

QY 784 TTCCAGGAGGAGATACACAGCTTTGGGTTTGGCTGTGGACGAGAAACCGCCCTGGAGTT 843
DB 766 TTCCAGGAGGAGATACACAGCTTTGGCTTTGGCTGTGGACGAGAAACCGCCCTGGAGTT 825
QY 844 TTCACTGCTGTGGCTACCTATGAGCATGGATACGGAGCAGGTGATGGTTTCAGAGCT 903
DB 826 TTCACTGCTGTGGCTACCTATGAGCATGGATACGGAGCAGGTGATGGTTTCAGAGCT 885
QY 904 GGGCTGCTCTTCCACCCAGCCCGAGAGAGACCCAGTTCAGAT 945
DB 886 GGGCTGCTCTTCCACCCAGCCCGAGAGAGACCCAGTTCAGAT 927
RESULT 4
ID ABQ75956
AC ABQ75956 standard; cDNA; 2681 BP.
XX ABQ75956;
XX 17-OCT-2002 (first entry)
XX Human PMMM encoding sequence Incyte ID 2751509CB1.
XX Human; PMMM; protein modification and maintenance molecule;
KW anticonvulsant; neuroprotective; nootropic; cytotstatic; antipsoriatic;
KW antiasthmatic; dermatological; antidiabetic; antiparkinsonian;
KW antianemic; antiinflammatory; antiulcer; antianginal; cardiant;
KW hepatotropic; osteopathic; antiemetic; antipyretic; virucide;
KW antibacterial; fungicide; gastrointestinal; antidiarrheic; laxative;
KW haemostatic; anti-Hiv; thrombolytic; anticoagulant; gynaecological;
KW cardiovascular; antiarteriosclerotic; hypotensive; vasotropic;
KW nephrotropic; antihypertensive; immunosuppressive; antiasthmatic;
KW nephrotoxic; antitumor; thyromimetic; antiarthritic; uropathic;
KW ophthalmological; antiparasytic; tranquiliser; vulnery; keratolytic;
KW auditory; antiseborrheic; antidepressant; neuroleptic; antiinfertility;
KW antihelmintic; prozoacide; Crohn's disease; hypertension; autoimmune;
KW inflammatory; anaemia; cell proliferative; developmental; epithelial;
KW scabies; neurological; Alzheimer's disease; reproductive;
KW ectopic pregnancy; gene therapy; vaccine; disorder; prostatic; gene; ss.
XX Homo sapiens.
OS Homo sapiens.
XX Key Location/Qualifiers
FH 1. .2457
FT /*tag= a
FT /product= "protein modification and maintenance molecule"
XX WC020246383-A2.
PN 13-JUN-2002.
XX 05-DEC-2001; 2001WO-US046964.
XX 08-DEC-2000; 2000US-0254399P.
PR 21-DEC-2000; 2000US-0257803P.
PR 05-JAN-2001; 2001US-0260110P.
PR 19-JAN-2001; 2001US-0262851P.
PR 25-JAN-2001; 2001US-0264623P.
XX (INCY-) INCYTE GENOMICS INC.
PA Yue H, Azimzai Y, Kallick DA, Baughn MR, Griffin JA, Swarnakar A;
XX Lal PG, Wadia NK, Hafalia AJA, Gandhi AR, Au-Young J, Elliott VS;
PI Ramkumar J, Thangavelu K, Lu Y, Warren BA, Lu DAM, Lee EA;
PI Tribouley CM, Arvizu C, Delegeane AM, Yao MG, Khan FA;
PI Sanjanwala MM;
XX WPI; 2002-519664/55.
DR P-PSDB; ABB98135.
XX New isolated Protein Modification and Maintenance polypeptides, useful
PT for diagnosis, and treatment of e.g. gastrointestinal disorders.
XX

PS Claim 5; Page 196; 200pp; English.
XX The invention relates to an isolated protein Modification and Maintenance
CC (PMM) polypeptide. Polypeptides of the invention may be used in the
CC diagnosis, treatment and prevention of disorders associated with
CC decreased expression or activity of PMMM. These include gastrointestinal
CC disorders (e.g. Crohn's disease), cardiovascular disorders (e.g.
CC hypertension), autoimmune/inflammatory disorders (e.g. anaemia), cell
CC proliferative disorders, developmental disorders, epithelial disorders
CC (e.g. scabies), neurological disorders (e.g. Alzheimer's disease)
CC reproductive disorders (e.g. ectopic pregnancy), and in gene therapy or a
CC vaccine for such diseases. They may also be used in the assessment of the
CC effects of exogenous compound on the expression of nucleic acid and amino
CC acid sequences of protein modification and maintenance molecules. The
CC current sequence represents a human PMMM encoding sequence of the
CC invention, encoding a polypeptide which has been found to have homology
CC with rat prostasin
XX
SQ Sequence 2681 BP; 431 A; 919 C; 861 G; 470 T; 0 U; 0 Other;
Query Match 74.0%; Score 815.6; DB 6; Length 2681;
Best Local Similarity 99.5%; Pred. No. 1.6e-134;
Matches 818; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
QY 124 CGTACTGCGGCGCCCTGAGCCCTGCGCCCGCATCTGTGGGGGGTCAACCGCAGCGG 183
DB 106 CTGGACTGCGGCGCCCTGAGCCCTGCGCCCGCATCTGTGGGGGGTCAACCGCAGCGG 165
QY 184 GGCACCTGCGCTTGGCAAGTGAAGCTGCACCATGAGGTGGCCACATCTGCGGGGGCTCC 243
DB 166 GGCACCTGCGCTTGGCAAGTGAAGCTGCACCATGAGGTGGCCACATCTGCGGGGGCTCC 225
QY 244 CTCATCGCCCGCCCTCTCGGTCCTCTCCGTCCTCTCTCTCATGCAATGGAGCTTG 303
DB 226 CTCATCGCCCGCCCTCTCGGTCCTCTCTCTCATGCAATGGAGCTTG 285
QY 304 GAGCCCGCGCGAGTGTGGTCTGCTGCGCTTGCATCTCCAGGAGGGGCGCCCTGGAC 363
DB 286 GAGCCCGCGCGAGTGTGGTCTGCTGCGCTTGCATCTCCAGGAGGGGCGCCCTGGAC 345
QY 364 GGGCGGCACACCGCGCAGTGGCCCGCATCTGCTGGCGCCCACTACAGCCAAAGTGGAG 423
DB 346 GGGCGGCACACCGCGCAGTGGCCCGCATCTGCTGGCGCCCACTACAGCCAAAGTGGAG 405
QY 424 CTGGCGCGCGACCTGCGCCCTGCTGCGCTTGCATCTCCAGGAGGGGCGCCCTGGGCTG 483
DB 406 CTGGCGCGCGACCTGCGCCCTGCTGCGCTTGCATCTCCAGGAGGGGCGCCCTGGGCTG 465
QY 484 TGGCCTGTCTGCTGCGCCCGCGCTTGCATCTCCAGGAGGGGCGCCCTGGGCTG 543
DB 466 TGGCCTGTCTGCTGCGCCCGCGCTTGCATCTCCAGGAGGGGCGCCCTGGGCTG 525
QY 544 ACCGGCTGGGAGAGCGTCCAGGAGGAGATCTCTGCTCTCTCCCTGGGTGTCTACAGGAA 603
DB 526 ACCGGCTGGGAGAGCGTCCAGGAGGAGATCTCTGCTCTCTCCCTGGGTGTCTACAGGAA 585
QY 604 GTGGAGCTTAAGGCTGTGCGCGAGGCGACCTGTCAATGTCTTACAGCAGCCCGGTGCC 663
DB 586 GTGGAGCTTAAGGCTGTGCGCGAGGCGACCTGTCAATGTCTTACAGCAGCCCGGTGCC 645
QY 664 TTCACCTCACTCTCCAGATATTTGCCAGGAGTGTGTGCTGGCTACCCAGAGGGCGCG 723
DB 646 TTCACCTCACTCTCCAGATATTTGCCAGGAGTGTGTGCTGGCTACCCAGAGGGCGCG 705
QY 724 AGGAGCACCTGCGCAGGGTGACCTCTGGGGGGCGCCCTGTGTGTGAGGAGAGCGCGCTGG 783
DB 706 AGGAGCACCTGCGCAGGGTGACCTCTGGGGGGCGCCCTGTGTGTGAGGAGAGCGCGCTGG 765
QY 784 TTCAGGCGAGGAATACACAGCTTTGGCTTGGCTGTGGAGCGAGAAACCGCCCTGGAGTT 843
DB 766 TTCAGGCGAGGAATACACAGCTTTGGCTTGGCTGTGGAGCGAGAAACCGCCCTGGAGTT 825
QY 844 TTCACCTGTCTGGCTACCTATGAGGATGATGAGGAGCAGGTGATGGTTTCAGAGCT 903

826 TTCACTGCTGGGTACCTATGAGCATGGATCGGAGCAGGTGCGTTGAGAGCCT 885
 904 GGGGCTGCTTCCACCCAGCCCGAGAGACCCAGTCAGAT 945
 886 GGGGCTGCTTCCACCCAGCCCGAGAGACCCAGTCAGAT 927

RESULT 5
 AAK94500
 ID AAK94500 standard; cDNA; 2810 BP.
 AC AAK94500;
 XX
 XX
 DT 06-NOV-2001 (first entry)
 XX
 XX Human full-length cDNA, SEQ ID NO: 3346.
 XX
 XX Human; full length cDNA; cDNA synthesis; oligo-capping; ss.
 KW
 XX Homo sapiens.
 OS
 XX EP1130094-A2.
 FN
 XX PD 05-SEP-2001.
 XX
 XX PF 07-JUL-2000; 2000EP-00114089.
 XX
 XX PR 08-JUL-1999; 99JP-00194486.
 PR 11-JAN-2000; 2000JP-00118774.
 PR 02-MAY-2000; 2000JP-00183765.
 XX
 XX PA (HELI-) HELIX RES INST.
 XX
 PI Ota T, Nishikawa T, Isogai T, Hayashi K, Ishii S, Kawai Y;
 PI Wakamatsu A, Sugiyama T, Nagai K, Kojima S, Otsuki T, Koga H;
 XX
 DR WPI; 2001-524255/58.
 P-PSDB; AM93568.
 XX
 PT 830 Primers useful for synthesizing full length cDNA clones and their use
 in genetic manipulation.
 XX
 PS Claim 8; SEQ ID NO 3346; 1380pp + Sequence Listing; English.
 XX
 CC The invention relates to primers for synthesising full length cDNA
 clones. 830 cDNA molecules encoding a human protein have been isolated
 and nucleotide sequences of 5'- and 3'-ends of the cDNA molecules have
 been determined. Primers for synthesising the full length cDNA are useful
 for clarifying the function of the protein encoded by the cDNA. The full
 length clones were obtained by construction of full length enriched cDNA
 libraries that were synthesised by the oligo-capping method. The primers
 enable the production of the full length cDNA easily without any special
 methods. The present sequence is a full length human cDNA of the
 invention. Note: The sequence data for this patent did not form part of
 the printed specification, but was obtained in CD-ROM format directly
 from EPO
 CC
 XX
 SQ Sequence 2810 BP; 436 A; 964 C; 902 G; 508 T; 0 U; 0 Other;

Query Match 73.2%; Score 806.6; DB 4; Length 2810;
 Best Local Similarity 92.0%; Pred. No. 6.2e-133;
 Matches 862; Conservative 0; Mismatches 74; Indels 1; Gaps 1;

QY 10 TCCTGGCCATGCCAGAGGGGTCCTGGGSCCTGGGCGAGCTGGGGGTGGGCCAAT 69
 23 TCATAGCGCCATGGCCCGGACCTGCTCTCCCTTGTGATGTTGTATCATGCTCCATC 82
 70 TCCTGACTCATPACTCACTTACGGGTTGGTCCGTCGGGACCCCGTACGGGCCCCCGGTAC 129
 83 CCAGGAGCCTTCAGGACTCAGTCTCAGTCTCAGTCTACCCAGGAGAACCTGAAGATCTGGAC 142
 130 TGGCGGGGCGCCTGAGCCCTCGGCCCGCATGCT-GGGGGGGCTCAACCGCGAGCCGGGCAC 188

143 TGGGGGGCCCTGAGCCCTCGGCCCGCATGTTGGGGGGGCTCAAAAGCGGAGCGGGGCAC 202
 189 CTGGCCTTGGCAAGTAGAGCTGACCATGAGAGTGGCCACATCTGGGGGGCTCCCTCAT 248
 203 CTGGCCTTGGCAAGTAGAGCTGACCATGAGAGTGGCCACATCTGGGGGGCTCCCTCAT 262
 249 CGGCCCTCTCTGGGTCTCTCCGCTGCTCACTGCTTTTCATGACGAATGGAGCTTGGAGCC 308
 263 CGGCCCTCTCTGGGTCTCTCCGCTGCTCACTGCTTTTCATGACGAATGGAGCTTGGAGCC 322
 309 CGGGCCAGTGGTGGTACTGCTGGGCGTGCATCCAGGACGGGCCCCCTGAGCGCGC 368
 323 CGGGCCAGTGGTGGTACTGCTGGGCGTGCATCCAGGACGGGCCCCCTGAGCGCGC 382
 369 GCACACCGCGGAGTGGCGCCCATCGTGGTGGCGGCCAACTACAGCCAAAGTGGAGCTGGG 428
 383 GCACACCGCGGAGTGGCGCCCATCGTGGTGGCGGCCAACTACAGCCAAAGTGGAGCTGGG 442
 429 CGCGACCTGGCCCTGCTGGGCGCTGAGCCCTACCGCCAGGCTGGGGCCCCCGCGTGTGGCC 488
 443 CGCGACCTGGCCCTGCTGGGCGCTGAGCCCTACCGCCAGGCTGGGGCCCCCGCGTGTGGCC 502
 489 TGTGTGCTGGCCCGCGCCCTCACAACCGCTTGTGTCACGGCACCGCCCTGCTGGGCGACCGG 548
 503 TGTGTGCTGGCCCGCGCCCTCACAACCGCTTGTGTCACGGCACCGCCCTGCTGGGCGACCGG 562
 549 CTGGGGAGAGCTCCAGGAGGAGATCTCTGCTCTCTCCCTGGGTGCTACAGGAAGTGGGA 608
 563 CTGGGGAGAGCTCCAGGAGGAGATCTCTGCTCTCTCCCTGGGTGCTACAGGAAGTGGGA 622
 609 GCTAAGGCTGCTGGGCGAGGCCACTGTCAATGTCTCTACAGCAGCCCGCTCCCTTCAA 668
 623 GCTAAGGCTGCTGGGCGAGGCCACTGTCAATGTCTCTACAGCAGCCCGCTCCCTTCAA 682
 569 CTTCACTCTCCAGATATTCAGGAGTGTGCTGTGGTGTACCCAGAGGCGCGCAGGGA 728
 683 CTTCACTCTCCAGATATTCAGGAGTGTGCTGTGGTGTACCCAGAGGCGCGCAGGGA 742
 729 CACTGCGCAGGAGTGTGCTGTGGGCGCGCCCTGCTGTGAGGAAGCGCGCCCTGTTCCA 788
 743 CACTGCGCAGGAGTGTGCTGTGGGCGCGCCCTGCTGTGAGGAAGCGCGCCCTGTTCCA 802
 789 GCGAGGAATACCGAGCTTGGGTTGGCTGTGAGCGGAGAAACCGCCCTGAGATTTCAC 848
 803 GCGAGGAATACCGAGCTTGGGTTGGCTGTGAGCGGAGAAACCGCCCTGAGATTTCAC 862
 849 TGCTGTGGTACCTATGAGGATGATACGAGGAGCAGGTGATGGGTTTCAGAGCCTTGGGCC 908
 863 TGCTGTGGTACCTATGAGGATGATACGAGGAGCAGGTGATGGGTTTCAGAGCCTTGGGCC 922
 909 TGCTTTTCCACCCAGCCCGAGAGACCCAGTTCAGAT 945
 923 TGCTTTTCCACCCAGCCCGAGAGACCCAGTTCAGAT 959

RESULT 6
 ABK13565/c
 ID ABK13565 standard; cDNA; 537 BP.
 XX
 AC ABK13565;
 XX
 DT 08-MAY-2002 (first entry)
 XX
 XX Human prostaasin-like serine protease cDNA #1.
 DE Human; prostaasin-like serine protease; cytotstatic; antiatherosclerotic;
 KW virucide; osteopathic; antinflammatory; vasotropic; neuroprotective;
 KW trypsin-like; mekasin; autoimmune lesion; atherosclerosis; gene;
 KW renal failure; malignant cell; tumour angiogenesis; enzyme; inflammation;
 KW chronic obstructive pulmonary disease; COPD; restenosis; atherosclerosis;
 KW neurodegenerative disease; prion protein; infection; amyloid plaque;
 KW Genstmann-Straussler Syndrome; viral infection; Scrapie;

Creutzfeldt-Jakob disease; metastatic cancer; lipid accumulation; osteoporosis; Paget's disease; ss.

Homo sapiens.

Key Location/Qualifiers
FH complement (3. .536)
FT /*tag= a
FT /product= "Prostasin-like enzyme"
ET /partial
ET /note= "No start or stop codons shown"

WO200198467-A2.

27-DEC-2001.

22-JUN-2001; 2001WO-EP007117.

23-JUN-2000; 2000US-0213588P.

20-MAR-2000; 2001US-0276909P.

(FARB) BAYER AG.

Xiao Y, Morozov V;

WPI; 2002-114576/15.

P-PSDB; AAU75082.

Novel human prostasin-like serine protease polypeptide and polynucleotide which can be regulated for treating metastasis of malignant cells, inflammation, atherosclerosis, neurodegenerative disease and infections.

Claim 1; Fig 1; 111pp; English.

This invention comprises the cDNA and protein sequences of an isolated prostasin-like serine protease and reagents and methods for regulating the human prostasin-like enzyme activity. Prostasin is a trypsin-like serine protease purified from human seminal fluid. An antibody specific for prostasin-like serine protease is useful for immunodetection and diagnosis of micro-metastases, autoimmune lesions and renal failure in biopsy specimens, plasma samples and body fluids. The antibody may be used to modulate enzyme activity in a disease, such as metastasis of malignant cells, tumour angiogenesis, inflammation, chronic obstructive pulmonary disease (COPD), atherosclerosis, neurodegenerative disease (e.g. prion protein amyloid plaques of Gerstmann-Strausler Syndrome, Creutzfeldt-Jakob disease, Scrapie) or infection, particularly viral infection. The human prostasin-like serine protease gene provides a therapeutic target of decreasing human prostasin-like serine protease activity, in particular for treating or preventing metastatic cancer. The agonists and antagonists of the nucleotide sequence may be used to mimic, augment and inhibit the enzyme activity which may be useful to treat osteoporosis, Paget's disease and degradation of bone implants particularly dental implants. Altered levels of human prostasin-like serine protease activity inhibits both smooth muscle cell proliferation and lipid accumulation and inhibit the progression of restenosis and atherosclerosis. The nucleic acid sequence is also useful in diagnostic assays for detecting diseases and abnormalities or susceptibility to diseases related to the presence of mutations in nucleic acid sequences which encode the enzyme. The present sequence represents the human prostasin-like serine protease #1 nucleotide sequence of the invention

Sequence 537 BP; 92 A; 176 C; 197 G; 72 T; 0 U; 0 Other;

Query Match 48.6%; Score 535.4; DB 6; Length 537;
Best Local Similarity 99.8%; Pred. No. 3e-85;
Matches 536; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

243 CCTCATCGCCCTCTCGGTCTCTCCGTCGTCTACTGTTTCATGACGAATGGAGCGTT 302
|||||
537 CCTCATCGCCCTCTCGGTCTCTCCGTCGTCTACTGTTTCATGACGAATGGAGCGTT 478
|||||
303 GGAGCCCGCGCGGAGTGTGCGTACTGTGGCGTGCATCCAGAGCGGCCCTCGGA 362
|||||

Db 477 GGAGCCCGCGCGGAGTGTGCGTACTGCTGGGCGTGCACATCCACAGACGGGCCCTTGA 418
QY 363 CGGCGCGCACACCGCGCAGTGGCGGCATCGTGGTCGGGCCAACTACAGCCAAAGTGGA 422
Db 417 CGGCGCGCACACCGCGCAGTGGCGGCATCGTGGTCGGGCCAACTACAGCCAAAGTGGA 358
QY 423 GCTGGGCGCGGACCTGCGCCCTGCTGCGCCCTGCGCCCTGCGCCCTGCGCCCTGCGCCCT 482
Db 357 GCTGGGCGCGGACCTGCGCCCTGCTGCGCCCTGCGCCCTGCGCCCTGCGCCCTGCGCCCT 298
QY 483 GTGGCTGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 542
Db 297 GTGGCTGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 238
QY 543 CACCGGCTGGGCGAGACGTCACAGGAGGAGATTCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 602
Db 237 CACCGGCTGGGCGAGACGTCACAGGAGGAGATTCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 178
QY 603 AGTGGAGCTAAGGCTGTGGGCGAGGCGACCTGTCAATGTCTTACAGCCAGCCCGGTC 662
Db 177 AGTGGAGCTAAGGCTGTGGGCGAGGCGACCTGTCAATGTCTTACAGCCAGCCCGGTC 118
QY 663 CTTCAACCTCACTCTCCAGATATTGCGGAGGATGCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 722
Db 117 CTTCAACCTCACTCTCCAGATATTGCGGAGGATGCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 58
QY 723 CAGGACACCTGCGCAGGCTGACTCTGGGGGCGCCCTGCTGTGTGAGGAAGCGGCGCG 779
Db 57 CAGGACACCTGCGCAGGCTGACTCTGGGGGCGCCCTGCTGTGTGAGGAAGCGGCGCG 1

RESULT 7

AAK92190

ID AAK92190 standard; cDNA; 670 BP.

XX AC AAK92190;

XX DT 06-NOV-2001 (first entry)

XX DE Human cDNA 5'-end sequence, SEQ ID NO: 650.

XX KW Human; full length cDNA; cDNA synthesis; oligo-capping; ss.

XX OS Homo sapiens.

XX PN EP1130094-A2.

XX PD 05-SEP-2001.

XX PF 07-JUL-2000; 2000EP-00114089.

XX PR 08-JUL-1999; 99JP-00194486.

XX PR 11-JAN-2000; 2000JP-00118774.

XX PR 02-MAY-2000; 2000JP-00183765.

XX PA (HELI-) HELIX RES INST.

XX PI Ota T, Nishikawa T, Isogai T, Hayashi K, Ishii S, Kawai Y;

XX PI Wakamatsu A, Sugiyama T, Nagai K, Kojima S, Otsuki T, Koga H;

XX XX WPI; 2001-524255/58.

XX DR 830 Primers useful for synthesizing full length cDNA clones and their use in genetic manipulation.

XX PT Claim 2; SEQ ID NO 650; 1380pp + Sequence Listing; English.

XX CC The invention relates to primers for synthesizing full length cDNA clones. 830 cDNA molecules encoding a human protein have been isolated and nucleotide sequences of 5'- and 3'-ends of the cDNA molecules have been determined. Primers for synthesizing the full length cDNA are useful for clarifying the function of the protein encoded by the cDNA. The full length clones were obtained by construction of full length enriched cDNA

CC libraries that were synthesised by the oligo-capping method. The primers
CC enable the production of the full length cDNA easily without any special
CC methods. The present sequence is the nucleotide sequence of the 5'-end of
CC a cDNA provided in the invention. Note: The sequence data for this patent
CC did not form part of the printed specification, but was obtained in CD-
CC ROM format directly from EPO

XX SQ Sequence 670 BP; 97 A; 245 C; 205 G; 118 T; 0 U; 5 Other;

Query Match 42.9%; Score 472.8; DB 4; Length 670;
Best Local Similarity 86.9%; Pred. No. 2.9e-74;
Matches 563; Conservative 0; Mismatches 81; Indels 4; Gaps 4;
QY 10 TCCTGGCCATGCCCCAGAGAGGGGTCTGGGCTGGGCGAGTGGGGCTGTGGCCAT 69
Db 23 TCTAGCCCATGGCCCCGGACCTGCTCTCCCTTGTGATGTCATCAGTCCCATC 82
QY 70 TCTGACTCATACTCATCTTACGGTGTGGTCCGACCCGCTAGGGCCCCCGCTAC 129
Db 83 CCAGGAGCCTTCCAGGACTCAGCTCTCAGTCTACCCAGGAGAACTGAGATCTGGAC 142
QY 130 TGGGGGCGCTGAGCCCTGGCCCGCATCGT-AGGGGGGTCAAACCGCAGCGGGCAC 188
Db 143 TGGGGGCGCTGAGCCCTGGCCCGCATCGTGGGGGGGTCAAACCGCAGCGGGCAC 202
QY 189 CTGSCCTTGGCAATGAGCTGACCATGGAGTGGCCACATCTGCGGGGCTCCCTCAT 248
Db 203 CTGGCCCTTGGCAATGAGCTGACCATGGAGTGGCCACATCTGCGGGGCTCCCTCAT 262
QY 249 CGCCCTCTCTGGGCTCTCCGCTGCTCACTGTTTCATGACGAATGGAGCTTGGAGCC 308
Db 263 CGCCCTCTCTGGGCTCTCCGCTGCTCACTGTTTCATGACGAATGGAGCTTGGAGCC 322
QY 309 CGGGCGCAGTGTGCGTACTGCTGGGCGTGCATCTCCAGGACGGGGCCCTGGACGGCC 368
Db 323 CGGGCGCAGTGTGCGTACTGCTGGGCGTGCATCTCCAGGACGGGGCCCTGGACGGCC 382
QY 369 GCACACCGCGCAGTGGCCGCTGCGGCTGCGGCGCAACTACAGCCAAAGTGGAGCTGGG 428
Db 383 GCACACCGCGCAGTGGCCGCTGCGGCTGCGGCGCAACTACAGCCAAAGTGGAGCTGGG 442
QY 429 CGCCGAGCTGCGCTGCTGGGCTGACCCGCTCAACCGCAGCTGGGCCCCCGGTGGCC 488
Db 443 CGCGGACTGGCCCTGTGGCTGCGCTCAACCGCTCAACCGCAGCTGGGCCCCCGGTGGCC 502
QY 489 TGTCTGCTGCGCCGCGCTCAGACCGCTTGTGCGACGGCACCGCTGCTGGGCGCACCGG 548
Db 503 TGTCTGCTGCGCCGCGCTCAGACCGCTTGTGCGACGGCACCGCTGCTGGGCGCACCGG 562
QY 549 CTGGGGAGAGCTGCA-GGAGGAGATCTCTGCTCTCCCTCTCCCTGGGTGTAC-AGGAAAGTG 606
Db 563 CTGGGGAGAGCTGCA-GGAGGAGGCAATTTCTTGTCTCTCCCTGGGTGTAC-AGGAAAGTG 622
QY 607 GAGCTAAGGCTGTGGGGG-AGGCCACTGTGCAATGTCTTACAGCA 653
Db 623 GAGCTAAGGCTGTGGGGGAGGCAACTGTGTCAATGTCTTACANNCA 670

RESULT 8

AAK93456

ID AAK93456 standard; cDNA; 670 BP.

XX AC

XX AC

XX AC

XX AC

DT 06-NOV-2001 (first entry)

XX XX

XX XX

XX XX

XX KW

XX OS

XX XX

XX PN

Human cDNA clone representative sequence, SEQ ID NO: 1916.

Human; full length cDNA; cDNA synthesis; oligo-capping; ss.

Homo sapiens.

Homo sapiens.

Homo sapiens.

Homo sapiens.

Homo sapiens.

Homo sapiens.

XX PD 05-SEP-2001.
XX PF 07-JUL-2000; 2000EP-00114089.
XX PR 08-JUL-1999; 99JP-00194486.
XX PR 11-JAN-2000; 2000JP-00118774.
XX PR 02-MAY-2000; 2000JP-00183765.
XX (HELI-) HELIX RES INST.
PA Ota T, Nishikawa T, Isogai T, Hayashi K, Ishii S, Kawai Y;
PI Wakamatsu A, Sugiyama T, Nagai K, Kojima S, Otsuki T, Koga H;
XX WPI; 2001-524255/58.
XX 830 Primers useful for synthesizing full length cDNA clones and their use
XX in genetic manipulation.
XX Example 11; SEQ ID NO 1916; 1380pp + Sequence Listing; English.

XX The invention relates to primers for synthesising full length cDNA
XX clones. 830 cDNA molecules encoding a human protein have been isolated
XX and nucleotide sequences of 5'- and 3'-ends of the cDNA molecules have
XX been determined. Primers for synthesising the full length cDNA are useful
XX for clarifying the function of the protein encoded by the cDNA. The full
XX length clones were obtained by construction of full length enriched cDNA
XX libraries that were synthesised by the oligo-capping method. The primers
XX enable the production of the full length cDNA easily without any special
XX methods. The present sequence was used as the representative sequence
XX from a human clone which was used in homology searches to identify the
XX clone. Note: The sequence data for this patent did not form part of the
XX printed specification, but was obtained in CD-ROM format directly from
XX EPO

XX SQ Sequence 670 BP; 97 A; 245 C; 205 G; 118 T; 0 U; 5 Other;

Query Match 42.9%; Score 472.8; DB 4; Length 670;
Best Local Similarity 86.9%; Pred. No. 2.9e-74;
Matches 563; Conservative 0; Mismatches 81; Indels 4; Gaps 4;

QY 10 TCCTGGCCATGCCCCAGAGAGGGGTCTGGGCTGGGCGAGTGGGGCTGTGGCCAT 69
Db 23 TCTAGCCCATGGCCCCGGACCTGCTCTCCCTTGTGATGTCATCAGTCCCATC 82
QY 70 TCTGACTCATACTCATCTTACGGTGTGGTCCGACCCGCTAGGGCCCCCGCTAC 129
Db 83 CCAGGAGCCTTCCAGGACTCAGCTCTCAGTCTTACCCAGGAGAACTGAGATCTGGAC 142
QY 130 TGGGGGCGCCCTGAGCCCTGGCCCGCATCGT-AGGGGGGTCAAACCGCAGCGGGCAC 188
Db 143 TGGGGGCGCCCTGAGCCCTGGCCCGCATCGTGGGGGGGTCAAACCGCAGCGGGCAC 202
QY 189 CTGSCCTTGGCAATGAGCTGACCATGGAGTGGCCACATCTGCGGGGCTCCCTCAT 248
Db 203 CTGGCCCTTGGCAATGAGCTGACCATGGAGTGGCCACATCTGCGGGGCTCCCTCAT 262
QY 249 CGCCCTCTCTGGGCTCTCCGCTGCTCACTGTTTCATGACGAATGGAGCTTGGAGCC 308
Db 263 CGCCCTCTCTGGGCTCTCCGCTGCTCACTGTTTCATGACGAATGGAGCTTGGAGCC 322
QY 309 CGGGCGCAGTGTGCGTACTGCTGGGCGTGCATCTCCAGGACGGGGCCCTGGACGGCC 368
Db 323 CGGGCGCAGTGTGCGTACTGCTGGGCGTGCATCTCCAGGACGGGGCCCTGGACGGCC 382
QY 369 GCACACCGCGCAGTGGCCGCTGCGGCTGCGGCGCAACTACAGCCAAAGTGGAGCTGGG 428
Db 383 GCACACCGCGCAGTGGCCGCTGCGGCTGCGGCGCAACTACAGCCAAAGTGGAGCTGGG 442
QY 429 CGCCGAGCTGCGCTGCTGGGCTGACCCGCTCAACCGCAGCTGGGCCCCCGGTGGCC 488
Db 443 CGCGGACTGGCCCTGTGGCTGCGCTCAACCGCTCAACCGCAGCTGGGCCCCCGGTGGCC 502


```
Db 308 CCGGTGAGTGGCCCTGCGAGTTCAGATACCACTATGAAGGGTCCATGTGTGGTGGC 367
Qy 241 TCCCTATCGCCCTCTCTGGTCTCTCCGCTGCTCACTGTTCATGACGAATGGGAGC 300
Db 368 TCTCTGCTGTCTGAGCAGTGGGTGTGTGAGCTGTCACTGTCTCCACAGGGA----- 420
Qy 301 TTGGAGCCGGCCGAGTGTGCTGCTACTGCTGGGGTGCACCTCCACAGAGGGGCCCTG 360
Db 421 --GCACCAAGGAAGCCATGAGGTCAAGCTTGGGGGCCACACAGTACCTTACTTCC 478
Qy 361 GACGGCGGCACACCCGGCGAGTGGCCGCCATCGTGTGGTGGCCGCAACTACAGCCAAAGTG 420
Db 479 GAGGAGCCCAAGGTGAGCACCTGGAAGGACATATCCCCACCCAGCTACTCTCCAGAG 538
Qy 421 GAGCTGGGGCGGAGCTGGCCCTGTGCGCTGCGCTCACCGCCAGCTGGGCCCCGCC 480
Db 539 GGCCTCCAGGGGCAATTTGCACTCTCTCAACTTCAGCAGACCCATCACTTCTCCGCTAC 598
Qy 481 GTGTGCGCTGTCTGCTGCCCCGCGCTCACACCGCTTGTGTCAGCGCACGCCCTGTGG 540
Db 599 ATCCGGCCCATCTGCTCTCCCTGACGCCAAGCGCTCTTCCCAAGCGCTCCACGTCACT 658
Qy 541 GCCACCGCTGGGAGACCTCCAGGAGGAGATCTCTCTCTCTCTCTCTCTCTCTCTCTCT 600
Db 659 GTCACTGGCTGGGTCACTGTGCCCCCTCAGTGAGCTCTCTGACGCCCAAGCACTGCAG 718
Qy 601 GAACTGGAGCTAAGCTGTGCGGAGGCGCACTGTCAATGTCTCTACAGCCAGCCCGGT 660
Db 719 CAACTCGAGGTGCTCTGTATCAGTCTGTGAGACGTGTAACTGTCTGTACAAATCAGAC 778
Qy 661 CCCTTCAACTCACTCTCCAGATATTGCGAGGATGCTGTGTGCTGGCTACCCAGAGGC 720
Db 779 AAGCTGAGGAGCGGACATTTGTCAGAGGACATGTGTGTGTGTGTGTGTGTGTGTGTGT 838
Qy 721 CGCAGGGAACCTGCCAGGGTCACTCTGGGGGGCCCCCTGTGTGTGTGTGTGTGTGTGTGT 780
Db 839 GSCAAGGAGCGCTGCCAGGTGACTCTGGGGGCCCACTCTCTGCTGCTGTGGAGGTCTC 898
Qy 781 TGGTTCAGGAGGATACCAAGCTTTGGCTGTGGTGTGGAGGAGAAACCGCCCTGGA 840
Db 899 TGGTACCTGACGGGCAATGTGAGCTGGGGAGATGCTGTGGGGCCCGCAACAGGCTGT 958
Qy 841 GTTTTCACTGCTGTGGCTACCTATGAGGCAATGATACGGGAGCAGGTGATGGTTTACAG 900
Db 959 GTGTACACTCTGCGCTTCCAGCTATGCTCTGTGATCCAAAGCAAGGTGACAGAACTCCAG 1018
Qy 901 CTGGGCGCTGCTTCCCAACCCAG 924
Db 1019 CCTCGTGTGGTGGCCCCCAACCCAG 1042

RESULT 12
AAF98720
ID AAF98720 standard; DNA; 1835 BP.
XX
AC AAF98720;
XX
DT 02-JUL-2001 (first entry)
XX
DE Human late stage ovarian tumour polynucleotide marker 28.
XX
KW Human; ovarian cancer; identification; detection; characterisation;
KW tumour; kinase; marker; cytostatic; antisense gene therapy; ds.
XX
OS Homo sapiens.
XX
XX WO200118542-A2.
XX
PD 15-MAR-2001.
XX
PF 01-SEP-2000; 2000WO-US024199.
XX
```

```
PR 03-SEP-1999; 99US-0152547P.
PR 16-MAR-2000; 2000US-0190347P.
PR 21-MAR-2000; 2000US-0191321P.
PR 31-MAY-2000; 2000US-0208382P.
PR 20-JUL-2000; 2000US-00220467.
XX (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
PA
XX Lee J, Thompsho P, Lillie J;
XX WPI; 2001-211428/21.
DR
XX
PT Detection, assessment, prevention and therapy of ovarian cancer,
PT comprises detecting changes in the expression of a variety of markers.
XX
PS Claim 1; Page 1186-1187; 1198pp; English.
XX
CC The present invention describes a method for assessing whether a patient
CC is afflicted with ovarian cancer by comparing: (1) the expression of a
CC marker (I) (see AAF98594 to AAF98730), in a patient sample; and (2) the
CC normal level of expression of (I) in a control non-ovarian cancer sample,
CC where a significant difference between the level of expression in (a) and
CC (b) is an indication that the patient is afflicted with ovarian cancer.
CC (I) have cytostatic activities and can be used in antisense gene therapy.
CC The method, compositions and kits from the present invention can be used
CC for: (1) assessing and treating ovarian cancer; (2) making isolated
CC hybridoma, which produces an antibody useful for ovarian cancer
CC assessment; and (3) inhibiting ovarian cancer in a patient. AAF98573 to
CC AAF98593 represent human kinase marker primers and probes which are used
CC in the exemplification of the present invention
XX
SQ Sequence 1835 BP; 309 A; 621 C; 527 G; 378 T; 0 U; 0 Other;
```

```
Query Match 21.1%; Score 232.4; DB 5; Length 1835;
Best Local Similarity 55.0%; Pred. No. 4.8e-32;
Matches 508; Conservative 0; Mismatches 401; Indels 15; Gaps 2;
Qy 1 GGGCCCTTGTCTCTGGCCATGCGCCAGAGGGGTCTCTGGGGCTTGGGCGAGCTGGGGGT 60
Db 212 GGGCCCTTGTCTCTGGCCATGCGCCAGAGGGGTCTCTGGGGCTTGGGCGAGCTGGGGGT 271
Qy 61 GTGGCCAAATTCGACTCATCTACTTACGGGTGTGTGCGTCCGGACCCGCTAGGGGC 120
Db 272 GTGGCCAAATTCGCTCTATCTTGGATTACTCTGGGTCTGGGGACAGGAGCGGAGCAHA 331
Qy 121 CCGCGTACTGCGGGCGCCCTGAGCCCTCGGCCCGCATCTGTGGGGGTCTCAACCGCAG 180
Db 332 GCTCCCTGGGTGTG-----GCCCGCCAGCACCCATCACAGTGGGAGCAGTGCAGTC 385
Qy 181 CCGGGCACCTGGCCCTTGGCAAGTGAAGCTGACCATGAGGTGGCCACATCTGCGGGGGC 240
Db 386 GCGGTGAGTGGCCCTGGCAGTCAAGTCACTATGAAGGGTCCATGTGTGTGGTGGC 445
Qy 241 TCCCTCATGCCCCCTCTCTGGTCTCTCTGGTCTCTCTGGTCTCTCTGATGACGATGGGAGC 300
Db 446 TCTCTGCTGTGAGCAGTGGGTGTGTGAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 498
Qy 301 TTGGAGCCCGCGCGAGTGGTGTGCTGTCTGTGGCGGTGCACTCCCGAGACGGGCCCTG 360
Db 499 --GCACCAAGGAAGCCATGAGGTCAAGTCTGGGGGCCACACAGTACCTTACTTCC 556
Qy 361 GACGGCGGCACACCCGGCGAGTGGCCGCCATCGTGGTGTGGGCGCAACTTACAGCCAAAGTG 420
Db 557 GAGGAGCCCAAGGTGAGCACCTGGAAGGACATATCCCGCCACCCAGCTACTCTCCAGGAG 616
Qy 421 GAGCTGGGGCGGAGCTGGCCCTGTGCGCTGTGGCTCACCGCCAGCTGGGCCCGGCC 480
Db 617 GGCCTCCAGGGGCAATTTGCACTCTCTCAACTTCAGCAGACCCATCACTTCTCCGCTAC 676
Qy 481 GTGTGCGCTGTGCTGCGCCCGCCCTCACACCGCTTGTGTCAGCGGACCCGCTGCTGG 540
Db 677 ATCCGGCCCATCTGCTGCGAGCCAGCGCTCTTCCCGCAAGGCTCCACTGCAGT 736
```

541 GCACCGCGCTGGGAGACGCTCCAGGAGGACAGATCCTCTGCTCTCCCTGGGTGCTACAG 600
 737 GTCACTGGCTGGGTCTATGTGGCCCCCTCACTGAGCTCTGAGCCCAAGCCACTGCAG 796
 601 GAAGTGGAGCTAAGGCTGTGGGGAGGCGCACTGTCTCAATGTCTCTACGCGCCCGGT 660
 797 CAATCGAGGTGCTCTGATCAGTGTGAGACGCTGTAACTGCTGTATCAACATCGAGCC 856
 661 CCCTTCAACCTCACTCTCCAGATATTCAGGATGCTGTGCTGCTACCCAGAGGGC 720
 857 AAGCTGAGGAGCGGCACTTTGTCTCAAGAGACATGTGTGTGCTATGTGGAGGG 916
 721 CGCAGGACACCTCCAGGCTGACTCTGGGGGGCCCTGTGTCTGTGAGGAGGCGCGC 780
 917 GGCAGGACGCTCCAGGCTGACTCTGGGGGGCCCTGTGTCTGTGAGGAGGCTC 976
 781 TGGTTCAGGAGGAATCACAGCTTTGGGTTTGGCTGTGACGAGGAACCGCCCTGGA 840
 977 TGGTACCTGACGGGCAATGTGAGCTGGGGAGATGCTGTGGGGCCCGCAAGGCTGGT 1036
 841 GTTTTCACTGCTGTGCTTACCTATGAGGCAATGATACGGGAGCAGGTGATGGTTTCAG 900
 1037 GTGTACACTCTGGGCTCGAGCTATGCTCTCTCTGATCCAAAGCAAGTGCAGAACTCC 1096
 901 CCTGGGCTGCTCTTTCCACCCAG 924
 1097 CCTGCTGTGGTGGCCCAACCCAG 1120
 RESULT 13
 ID ABZ35336
 XX ABZ35336 standard; cDNA; 3382 BP.
 AC ABZ35336;
 XX 05-FEB-2003 (first entry)
 DT Human gene expression profile polynucleotide SEQ ID NO 447.
 DE Human;
 KW Human; artery; endothelium; umbilical; vein; aorta; pulmonary artery;
 KW bronchial epithelium; prostate; muscle; lung fibroblast; osteoblast;
 KW tumour; microarray; genome mapping; antibiotic; antiviral; antifungal;
 KW gene expression; gene; ss.
 XX Homo sapiens.
 OS
 XX WO200274979-A2.
 PN
 XX 26-SEP-2002.
 PD
 XX 20-MAR-2002; 2002WO-US008456.
 PF
 XX 20-MAR-2001; 2001US-0276947P.
 PR
 XX (ORTH) ORTHO CLINICAL DIAGNOSTICS INC.
 XX Wan J, Wang Y;
 PI
 XX WPI; 2002-740862/80.
 DR
 XX New gene expression profile generated from primary, endothelial,
 PT epithelial, and muscle cell types, useful for identifying disease
 PT pathologies involving alterations of gene expression, e.g. cancer.
 XX
 XX Example 3; Page 595-496; 850pp; English.
 PS
 XX The invention relates to a gene expression profile comprising one or more
 CC genes (ABZ3489-ABZ35692) and generated from a cell type. The cell type
 CC is a coronary artery endothelium, umbilical artery or vein endothelium,
 CC aortic endothelium, dermal microvascular endothelium, pulmonary artery
 CC endothelium, myometrium microvascular endothelium, keratinocyte
 CC epithelium, bronchial epithelium, mammary epithelium, prostate
 CC epithelium, renal cortical epithelium, renal proximal tubule epithelium,

CC small airway epithelium, renal epithelium, umbilical artery smooth
 CC muscle, neonatal dermal fibroblast, pulmonary artery smooth muscle,
 CC dermal fibroblast, neural progenitor cells, skeletal muscle, astrocytes,
 CC aortic smooth muscle, mesangial cells, coronary artery smooth muscle,
 CC bronchial smooth muscle, uterine smooth muscle, lung fibroblast,
 CC osteoblasts or prostate stromal cell. The gene expression profile is used
 CC for determining the level of RNA expression for a sample, determining the
 CC phenotype of a cell and distinguishing cell types. The gene or a protein
 CC expression profile is useful in identifying disease pathologies involving
 CC alterations of gene expression. The assessment of expression profiles may
 CC provide meaningful information with respect to tumour type and stage,
 CC treatment methods, and prognosis. The gene or protein expression profile
 CC may also be used for creating microarrays. The microarray is useful for
 CC genetic and physical mapping of genomes, DNA sequencing, genetic or
 CC medical diagnosis, genotyping of organisms, confirming cell or tissue
 CC identifications and in identifying promising antibiotics, antiviral or
 CC antifungal agents
 XX
 SQ Sequence 3382 BP; 602 A; 1069 C; 959 G; 752 T; 0 U; 0 Other;
 Query Match 21.1%; Score 232.4; DB 6; Length 3382;
 Best Local Similarity 55.0%; Pred. No. 4.7e-32;
 Matches 508; Conservative 0; Mismatches 401; Indels 15; Gaps 2;
 QY 1 GGGCCCTTGTCTGGCCATGCGCCAGAGGGGGTCTGGGGGCTGGGCGAGCTGGGGGCT 60
 DB 212 GGGCCCTTGTCTGGCCATGCGCCAGAGGGGGTCTGGGGGCTGGGCGAGCTGGGGGCT 271
 QY 61 GTGGCCCAATTCGACTCATACTCATTAGGGTGTGGTCCGTCGAGACCCGCTAGGGGC 120
 DB 272 GTGGCCCAATTCGACTCATACTCATTAGGGTGTGGTCCGTCGAGACCCGCTAGGGGC 331
 QY 121 CCCCCTGACTCGCGGGCCCTCGAGCCCTCGCGCCGCTATCGTGGGGGCTCAAAAGCGCAG 180
 DB 332 GCTCCCTGGCGTGTG-----GCCCCCAAGCAGCATCACAGTGGCAGCAGTCAGTC 385
 QY 181 CCGGGCACCTTGGCTTGGCAAGTGAGCTGCAATGAGAGTGGCCCACTCTGGGGGGC 240
 DB 386 GCGGTCAGTGGCCCTGGCAGGTGAGCATCATATGAAGGCGTCCATGTGTGTGGTGGC 445
 QY 241 TCCTCATCGCCCTCTCGGTCTCTCGCTGCTCTCAGTGTTCATGAGATGGGACG 300
 DB 446 TCTCTCTGTGTGAGCAGTGGGTGTGTGTGAGTGTCTCAGTGTCTCCCGAGCGA----- 498
 QY 301 TTGAGCCCGCGCGCGAGTGTGTGGTGTCTGTGGCGTGTGCACTCCAGGAGCGGCCCTG 360
 DB 499 --GCACCACAAGGAAGCCTATGAGTCAAGCTGGGGGCCCAACAGCTAGACTCTCTACTCC 556
 QY 361 GACGGCGGCACACCCCGCAGTGGCGCCATGTGTGTGGCGCCCACTACAGCAGTCAGTG 420
 DB 557 GAGGACGCCAAGGTTCAGCACCCCTGAAGGACATATCCCCCAACCCAGCTACCTCCAGGAG 616
 QY 421 GAGCTGGCGCGCAGCTGGCCCTGCTGCGCTGCTGCGCTCACCCTGAGCTGGGGCCCGCC 480
 DB 617 GGCTCCAGGGGACATTCAGTCTCTCACTGAGCAGACCCATCACTCTCTCCGCTAC 676
 QY 481 GTGTGGCTGTGTCTGCTGCCCCCGCGCTTCACACCGCTTGTGTGACCGGCACCGCTGTGG 540
 DB 677 ATCGGCCCCATCTGCTCTCCCTCGCAGCCAAAGCCTCTTCCCAACAGCCCTCACTGCACT 736
 QY 541 GCGACCGGCTGGGAGAGCTCCAGGAGCAGATCTCTGCTCTCTCCCTGGGTGCTACAG 600
 DB 737 GTCACTGGCTGGGTGTATGTGGCCCCCTCAGTGTGAGCTCTCTGAGCCCAAGCCACTGAG 796
 QY 601 GAAGTGGAGCTAAGGCTGTGGGGAGGCGCCACCTGTCAATGTCTCTACAGCAGCCCGGT 660
 DB 797 CAATTCAGGCTGCTCTGTATCAGTGTGAGACGCTGTAACTGCTGTATCAACATCGAGCC 856
 QY 661 CCCTTCAACCTCACTCTCCAGATATTCAGGAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 720
 DB 857 AAGCTCGAGGAGCGCACTTTGTCTCAAGAGGACATGAGTGTGTGTGTGTGTGTGTGTGTGTGT 916
 QY 721 CGCAGGAGACCTGCCAGGAGTGTCTGTGGGGGGCCCTGTGTGTGTGTGTGTGTGTGTGTGTGT 780

Db 917 GGCAGAGCGCTGCGAGGTGATCTTGGGGGCCCACTCTCTGCTTGGAGGTC 976
Qy 781 TGGTCCAGGAGGAATCACAGCTTTGGGTTGGCTGTGGACGGAGAAACCGCCCTGGA 840
Db 977 TGGTACCTGACGGGCAATGTGAGCTGGGGAGATGCTGTGGGGCCCGCAACAGGCTTGGT 1036
Qy 841 GTTTTCACTGCTGGTACCTATGAGCATGAGTACGGGAGCAGGTGATGGTTTCAGAG 900
Db 1037 GTGTACACTCTGGGCTCAGTATGCTCTCTGGATCCAAAGCAAGGTGACAGAACTCCAG 1096
Qy 901 CCTGGGCTGCTTTCACCCAG 924
Db 1097 CCTGTGTGGTCCCAACCCAG 1120

RESULT 14
ABT31936
ID ABT31936 standard; DNA; 1733 BP.
XX AC ABT31936;
XX DT 01-MAY-2003 (first entry)
XX DE Human breast cancer / ovarian cancer related coding sequence #43.
XX KW Human; gene; ds; cytostatic; breast cancer; ovarian cancer.
XX OS Homo sapiens.
XX FN WC2003000012-A2.
XX PD 03-JAN-2003.
XX PF 21-JUN-2002; 2002MO-US019773.
XX PR 21-JUN-2001; 2001US-0300159P.
XX PR 27-JUN-2001; 2001US-0301351P.
XX XX (MILL-) MILLENNIUM PHARM INC.
XX PA Veiby OP;
XX PI WPI; 2003-267848/26.
XX DR P-PSDB; ABJ37067.
XX PS Determining the presence of breast cancer in an individual, involves using specific polynucleotide markers.
XX PT Disclosure; Page 192-193; 233pp; English.
XX CC The invention comprises a method for assessing whether a patient is afflicted with breast cancer or ovarian cancer. The method involves the use of specific DNA markers. The method of the invention is useful in the detection and treatment of ovarian and breast cancer. DNA sequences ABT31894 - ABT31949 encode human breast/ovarian cancer-related proteins
XX XX Sequence 1733 BP; 311 A; 578 C; 500 G; 344 T; 0 U; 0 Other;
SQ

Query Match 20.9%; Score 230.8; DB 7; Length 1733;
Best Local Similarity 54.9%; Pred. No. 9.2e-32;
Matches 507; Conservative 0; Mismatches 402; Indels 15; Gaps 2;

Qy 1 GGGCCCTTCTGCTGGCCATGCGCCAGAGGGGTCCTGGGGCCTGGGAGCTGGGGGCT 60
Db 81 GGGCCCTTCTGCTGGCCATGCGCCAGAGGGGTCCTGGGGCCTGGGAGCTGGGGGCT 140
Qy 61 GTGGCCAAATTCGACTCACTTACGGGTTGGTGCCTCCGGACCCCGTAGGGGC 120
Db 141 GTGGCCATTCCTCTATCTTGGATTACTCCGGTCAGGGACAGAGCGGAAGGGGCAGAA 200
Qy 121 CCCCCGTACTGGGGCGCCCTGAGCCCTGGGCCCGCATGTGGGGGGCTCAACCGCAG 180

Db 201 GCTCCCTGGGGTGTG-----GCCCCCAAGCAGCATCAAGGTGGCAGAGTGCAGTC 254
Qy 181 CCGGCACTTGGGCTTGGCAAGTGAGCTTGCACCAATGAGAGGTGGCCACATCTTGGGGGGC 240
Db 255 GCGGTCAGTGGGCTTGGCAGAGTCAGCATCACCTATGAAGGCGTCCATGTGTGTGGTGGC 314
Qy 241 TCCTCATGCCCCCTCTCTCGCTGCTCGCTGCTCAGTCTCAGTGTTCATGACGAATGGGAGC 300
Db 315 TCTCTGTGTGTGAGCAGTGGGTGCTGTGAGTGTCTCAGTGTCTCAGTGTCTTCCCGGGA----- 367
Qy 301 TTGAGCCCCCGGCGAGTGTGTGCTGCTGCGGCTGCACTCTCCAGAGCGGGCCCTG 360
Db 368 --GCACCAAGGAAGCTATGAGGTCAAGCTGGGGGCCACACAGCTAGACTCTCTACTCC 425
Qy 361 GACGGCGGCACACCCGCGCAGTGGCCGCTATGCTGTGTGTGGCCCACTACAGCAAGTG 420
Db 426 GAGGACGCCAAGTCAAGCTTCAAGGACATCATCCCCCACCCAGCTACCTCCAGGAG 485
Qy 421 GAGCTGGGCGCGACCTTGGGCTGCTGCGCTGCGCTCACCGGCGAGCTGGGGCCCGGCC 480
Db 486 GGCTCCCGAGGGGACATTCACCTCTCAACTCAGCAGACCCATCACTTCTCCGGTAC 545
Qy 481 GTGTGCTGTGTGCTGCTGCGGCGCTCACACGCTTGTGTGACGGCAGCGCTGTGTGG 540
Db 546 ATCGGGCCCATCTGCTTCCCTGAGCCAAAGCTCTTCCCAACGGGCTCCACTGCACT 605
Qy 541 GCCACCGCTGGGGAGAGCTCCAGAGGAGATCTCTGCTCTCTCCCTGGGTCTACAG 600
Db 606 GTCACTGGCTGGGGTCAATGTGGGCCCCCTCAGTGAGCTCTCTGACGCCCAAGCCACTG 665
Qy 601 GAAGTGAGCTAAGGCTGCTGGGCGAGGCCACCTGTCAATGTCTCTACAGCCAGCCCGGT 660
Db 666 CACTCGAGGTGCTCTGATCAGTCTGTGAGCGTGTAACTGCTGTACAACTCAGCGCC 725
Qy 661 CCCTTCAACTCTCTCCAGATATGCCAGGAGTGTGTGTGTGTGTGTGTGTGTGTGTGT 720
Db 726 AAGCCTGAGGAGCGCACTTTGTCCAAAGAGGACATGTGTGTGTGTGTGTGTGTGTGT 785
Qy 721 CGCAGGACACCTGCCAGGTGACTCTGGGGGGCCCTGGTCTGTGAGGAGCGGGCCG 780
Db 786 GGCAGGACGCTGCCAGGTGACTCTGGGGGGCCCACTCTCTGCTGTGTGTGTGTGTGT 845
Qy 781 TGGTTCCAGGAGGAATCACAGCTTTGGGTTTGGCTGTGACGGAGAAACCGCCCTGGA 840
Db 846 TGTATCTGACGGGCAATTTGTGAGCTGGGAGATGCTGTGGGGCCCGCAACAGGCTTGGT 905
Qy 841 GTTTTCACTGTGTGGTACCTATGAGGCATGATAGGGAGCAGGTGATGGTTTCAGAG 900
Db 906 GTGTACACTCTGGGCTCCAGCTATGCTCTCTGGATCCAAAGCAAGGTGACAGAACTCCAG 965
Qy 901 CCTGGGCTGCTTTCACCCAG 924
Db 966 CCTGTGTGTGGTCCCAACCCAG 989

RESULT 15
ABA94396
ID ABA94396 standard; cDNA; 944 BP.
XX AC ABA94396;
XX DT 26-MAR-2002 (first entry)
XX DE Human prostatic-like serine protease encoding cDNA.
XX KW Prostatic-like enzyme; human; prostatic-like serine protease; cytostatic; antiatherosclerotic; virucide; osteopathic; antiinflammatory; vasotropic; neuroprotective; gene therapy; antisense therapy; ss.
XX OS Homo sapiens.
XX FH Key Location/Qualifiers
FT CDS 1..819

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 25, 2004, 08:20:46 ; Search time 2783 Seconds

(without alignments)
11824.694 Million cell updates/sec

Title: US-10-037-417-45

Perfect score: 1102

Sequence: 1 gggccctgtctgggcat.....ggggtttctgtgggctcc 1102

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 27513289 seqs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:*

1: em_estba.*

2: em_esthum.*

3: em_estin.*

4: em_estmu.*

5: em_estov.*

6: em_estpl.*

7: em_estro.*

8: em_htc.*

9: gb_estci.*

10: gb_est2.*

11: gb_htc.*

12: gb_est3.*

13: gb_est4.*

14: gb_est5.*

15: em_estfun.*

16: em_estom.*

17: em_gss_hum.*

18: em_gss_inv.*

19: em_gss_pln.*

20: em_gss_vrt.*

21: em_gss_fun.*

22: em_gss_mam.*

23: em_gss_mus.*

24: em_gss_pro.*

25: em_gss_rod.*

26: em_gss_phg.*

27: em_gss_vrl.*

28: gb_gsl.*

29: gb_gsl2.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Match	Query Length	ID	Description
C 1	533.8	48.4	537	10	AW450407 UI-H-B13-
2	442.8	40.2	456	9	AI190509 qd49f07.x
3	214.2	19.4	1071	9	AL551470 AL551470
4	202	18.3	433	10	BB850564 BB850564

5	199.2	18.1	433	10	BB850607
6	194.6	17.7	1699	11	AK078696
7	191.8	17.4	1013	12	BM923713
8	191.2	17.4	433	10	BB850795
9	180	16.3	1629	11	AK010640
10	175	15.9	861	12	BI763558
11	174.8	15.9	1010	13	BY710051
12	171.8	15.6	429	10	BB849543
13	171.4	15.6	930	13	BX436299
14	169.6	15.4	249	10	BB859364
15	168.6	15.3	430	28	AZ252377
16	164	14.9	1150	11	AK080281
17	161.4	14.6	1188	11	BC034294
18	160	14.5	947	9	AL555870
19	158	14.3	388	13	BY234907
20	157.8	14.3	1050	11	AK06271
21	157.4	14.3	765	12	BG762809
22	157.4	14.3	1135	12	BM917234
23	156.4	14.2	946	9	AL578261
24	156.2	14.2	865	14	CD389322
25	155.8	14.1	3035	11	AK004939
26	155.6	14.1	365	13	BY235958
27	155.4	14.1	699	14	CB851297
28	154.6	14.0	369	13	BY236911
29	151.6	13.8	777	12	BI259237
30	150.4	13.6	1201	13	BX417595
31	150	13.6	888	14	CA489381
32	148.4	13.5	922	13	BQ690230
33	147.6	13.4	591	10	BE590187
34	144.6	13.1	909	14	CD050808
35	144	13.1	352	29	CE450243
36	143.4	13.0	1018	13	BY706352
37	143.4	13.0	1052	12	BI54641
38	143	13.0	748	12	BI768651
39	142.6	12.9	786	13	BX644781
40	142	12.9	619	12	BM794592
41	142	12.9	815	14	CB995955
42	141.8	12.9	351	13	BY236668
43	141.8	12.9	653	14	CF900000
44	141.4	12.8	646	14	CB054579
45	140.8	12.8	1323	11	AK014645

ALIGNMENTS

RESULT 1
AW450407/c
LOCUS
DEFINITION
UI-H-B13-akn-g-11-0-UI-s1 NCI CGAP_Sub5 Homo sapiens cDNA clone
IMAGE:2735037 3', mRNA sequence.
ACCESSION
AW450407
VERSION
AW450407.1 GI:6991183
KEYWORDS
EST.
SOURCE
Homo sapiens (human)
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
REFERENCE
1 (bases 1 to 537)
NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
AUTHORS
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
TITLE
Tumor Gene Index
JOURNAL
Unpublished (1997)
COMMENT
Contact: Robert Strausberg, Ph.D.
Email: cgaps-remail.nih.gov
Oligo-dt track not found, Not I site shown in beginning of sequence
is likely internal to the message. cDNA Library Preparation: M.B.
Soares Lab Clone distribution: NCI-CGAP clone distribution
information can be found through the I.M.A.G.E. Consortium/LLNL at:
www.bio.llnl.gov/bbrp/image/image.html
Seq primer: M3 Forward
POLYA=No.

FEATURES
Location/Qualifiers

AUTHORS Akimura,T., Arakawa,T., Carninci,P., Furuno,M., Sugahara,Y., Shibata,K., Hayatsu,N., Hiramoto,K., Hirooka,T., Hirozane,T., Imotani,K., Ishii,Y., Ito,M., Kawai,J., Kojima,Y., Konno,H., Kouda,M., Matsuyama,T., Nakamura,M., Nishi,K., Nomura,K., Numasaki,R., Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K., Sakazume,N., Sasaki,D., Sato,K., Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H., Tagawa,A., Takahashi,F., Takaku-Akahira,S., Tanaka,T., Tomaru,A., Toya,T., Watahiki,A., Yasunishi,A., Muramatsu,M. and Hayashizaki,Y.

TITLE RIKEN Encyclopedia of Mouse Full-length cDNAs (Akimura,T., et al. 2001)

JOURNAL Unpublished (2001)

COMMENT Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@gsc.riken.go.jp/
URL: http://genome.gsc.riken.go.jp/
Carninci,P., Shibata,Y., Hayatsu,N., Muramatsu,M. and Hayashizaki,Y.
Itoh,M., Konno,H., Okazaki,Y., Sugahara,Y., Shibata,K., K.,
Normalizaton and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. Genome Res. 10 (10), 1617-1630 (2000)
wagi,K., Fujiwaka,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsura,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer. Genome Res. 10 (11), 1757-1771 (2000)
Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P., Sugahara,Y. and Hayashizaki,Y.
Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
Please visit our web site (http://genome.gsc.riken.go.jp) for further details.
e mouse tissues.
Location/Qualifiers
1. .433
/organism="Mus musculus"
/mol_type="rRNA"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="F930108A18"
/tissue_type="inner ear"
/dev_stage="adult"
/clone_lib="RIKEN full-length enriched, adult inner ear"

FEATURES source

Query Match 18.3%; Score 202; DB 10; Length 433;
Best Local Similarity 81.0%; Pred. No. 5.9e-26;
Matches 235; Conservative 0; Mismatches 55; Indels 0; Gaps 0;

ORIGIN

84 ACTTTAGGGTTGGTCCGTCGGACCCCGCTAGGGGCCCGCCCGTACTGCGGGCGCCCTGA 143
141 ACTCAGCTGGTCAGCTCCTTACCAAGAGAAATTGAAGACCTAGACTGCGGGCGCCCTGA 200
144 GCCTCGGGCCGATCGTGGGGGGTCAAAACGGCAGCGCGGCACCTGGCTTGGCAAGT 203
201 GCCCTTTCCGATGTTGGGGGGCTCGACGCTGATCTGGCACTTGGCCATGGCAGT 260
204 GAGCCTGCACCATGAGGTGGCCACATCTGCGGGGGCTCCCTATCGCCCGCTCTCTGGGT 263
261 GAGCCTGCATCAGGTGGGGGCCACATCTGCGGGGGCTCCCTATCGCCCGCTCTCTGGGT 320
264 CTTCTCGCTGCTCAGTGTTCATGACGATGGGACGTTGGAGCCCGCGCGAGTGTC 323
321 CCTCTCGCTGCTCAATGTTCTGTCGAAATGGAACCTTGGAGCCCGCGGAGAAATTGTC 380

QY 324 GGTACTGCTGGCGGTGCACCTCCAGGACGGGCCCTTGGACGGCGCGCAC 373

Db 381 AGTTCTGCTGGCGGTGCACCTCCAGGACGGGCCCTTGGAGGAGCGCAC 430

RESULT 5

LOCUS BB850607 433 bp mRNA linear EST 26-NOV-2001

DEFINITION BB850607 RIKEN full-length enriched, adult inner ear Mus musculus cDNA clone F930108E12 5', mRNA sequence.

ACCESSION BB850607

VERSION BB850607.1 GI:17092061

KEYWORDS EST.

SOURCE Mus musculus (house mouse)

ORGANISM Mus musculus

REFERENCE 1 (bases 1 to 433)
Akimura,T., Arakawa,T., Carninci,P., Furuno,M., Hanagaki,T., Hayatsu,N., Hiramoto,K., Hirooka,T., Hirozane,T., Imotani,K., Ishii,Y., Ito,M., Kawai,J., Kojima,Y., Konno,H., Kouda,M., Matsuyama,T., Nakamura,M., Nishi,K., Nomura,K., Numasaki,R., Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K., Sakazume,N., Sasaki,D., Sato,K., Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H., Tagawa,A., Takahashi,F., Takaku-Akahira,S., Tanaka,T., Tomaru,A., Toya,T., Watahiki,A., Yasunishi,A., Muramatsu,M. and Hayashizaki,Y.
RIKEN Encyclopedia of Mouse Full-length cDNAs (Akimura,T., et al. 2001)

TITLE Unpublished (2001)

JOURNAL Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@gsc.riken.go.jp/
URL: http://genome.gsc.riken.go.jp/
Carninci,P., Shibata,Y., Hayatsu,N., Muramatsu,M. and Hayashizaki,Y.
Itoh,M., Konno,H., Okazaki,Y., Sugahara,Y., Shibata,K.,
Normalizaton and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. Genome Res. 10 (10), 1617-1630 (2000)
wagi,K., Fujiwaka,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsura,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer. Genome Res. 10 (11), 1757-1771 (2000)
Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P., Sugahara,Y. and Hayashizaki,Y.
Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
Please visit our web site (http://genome.gsc.riken.go.jp) for further details.
e mouse tissues.
Location/Qualifiers
1. .433
/organism="Mus musculus"
/mol_type="rRNA"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="F930108E12"
/tissue_type="inner ear"
/dev_stage="adult"
/clone_lib="RIKEN full-length enriched, adult inner ear"

ORIGIN

Query Match 18.1%; Score 199.2; DB 10; Length 433;
Best Local Similarity 86.9%; Pred. No. 1.9e-25;

Matches	219;	Conservative	0;	Mismatches	33;	Indels	0;	Gaps	0;
QY	122	CCCCGTACTCGGGGGCCCTCGAGCCCTCGGGCCCGCATCGTGGGGGGCTCAAAACGGCGACG	181						
Db	179	CCTAGACCTGCGGCGCCCTGAGCCCTCTTCCGCACTTGTGGGGGCTCGGACGCTCATC	238						
QY	182	CGGGCACTTGGCCTTGGCAAGTGAAGCTGCACCATGAGAGTGGCCACATCTCGGGGGCT	241						
Db	239	CTGGAACCTGGCATGGCAGGTGAGCTGCATCAAGTGGGGGGCCACATCTCGGGGGCT	298						
QY	242	CCCTCATGCCCCCTCTCGGTGCTCTCGGTGCTCACATGTTTCATGACGAATGGGACGT	301						
Db	299	CCCTCATGCCCCCTCTCGGTGCTCTCGGTGCTCACATGTTTCATGACGAATGGGACGT	358						
QY	302	TGGAGCCCCGGCGGAGTGTGCTGTA	361						
Db	359	TGGAGCCGGGACGAATTGTGAGTTCTGCTGGCGTGCACATCCAGGACGGGCCCTGG	418						
QY	362	ACGGCGCGGACA	373						
Db	419	AAGGAGCGCAAA	430						
RESULT	6								
AK078696									
LOCUS									
DEFINITION									
ACCESSION									
VERSION									
KEYWORDS									
SOURCE									
ORGANISM									
REFERENCE									
AUTHORS									
TITLE									
JOURNAL									
MEDLINE									
PUBMED									
REFERENCE									
AUTHORS									
TITLE									
JOURNAL									
MEDLINE									
PUBMED									
REFERENCE									
AUTHORS									
TITLE									
JOURNAL									
MEDLINE									
PUBMED									
REFERENCE									
AUTHORS									
TITLE									
JOURNAL									
MEDLINE									
PUBMED									
REFERENCE									
AUTHORS									
TITLE									
JOURNAL									
MEDLINE									
PUBMED									
REFERENCE									
AUTHORS									
TITLE									
JOURNAL									
MEDLINE									
PUBMED									
REFERENCE									
AUTHORS									
TITLE									
JOURNAL									
MEDLINE									
PUBMED									
REFERENCE									
AUTHORS									
TITLE									
JOURNAL									
MEDLINE									
PUBMED									
REFERENCE									
AUTHORS									
TITLE									
JOURNAL									
MEDLINE									
PUBMED									
REFERENCE									
AUTHORS									
TITLE									
JOURNAL									
MEDLINE									
PUBMED									
REFERENCE									
AUTHORS									
TITLE									
JOURNAL									
MEDLINE									
PUBMED									
REFERENCE									
AUTHORS									
TITLE									
JOURNAL									
MEDLINE									
PUBMED									
REFERENCE									
AUTHORS									
TITLE									
JOURNAL									
MEDLINE									
PUBMED									
REFERENCE									
AUTHORS									
TITLE									
JOURNAL									
MEDLINE									
PUBMED									
REFERENCE									
AUTHORS									
TITLE									
JOURNAL									
MEDLINE									
PUBMED									
REFERENCE									
AUTHORS									
TITLE									
JOURNAL									
MEDLINE									
PUBMED									
REFERENCE									
AUTHORS									
TITLE									
JOURNAL									
MEDLINE									
PUBMED									
REFERENCE									
AUTHORS									
TITLE									
JOURNAL									
MEDLINE									
PUBMED									
REFERENCE									
AUTHORS									
TITLE									
JOURNAL									
MEDLINE									
PUBMED									
REFERENCE									
AUTHORS									
TITLE									
JOURNAL									
MEDLINE									
PUBMED									
REFERENCE									
AUTHORS									
TITLE									
JOURNAL									
MEDLINE									
PUBMED									
REFERENCE									
AUTHORS									
TITLE									
JOURNAL									
MEDLINE									
PUBMED									
REFERENCE									
AUTHORS									
TITLE									
JOURNAL									
MEDLINE									
PUBMED									
REFERENCE									
AUTHORS									
TITLE									
JOURNAL									
MEDLINE									
PUBMED									
REFERENCE									
AUTHORS									
TITLE									
JOURNAL									
MEDLINE									
PUBMED									
REFERENCE									
AUTHORS									
TITLE									
JOURNAL									
MEDLINE									
PUBMED									
REFERENCE									
AUTHORS									
TITLE									
JOURNAL									
MEDLINE									
PUBMED									
REFERENCE									
AUTHORS									
TITLE									
JOURNAL									
MEDLINE									
PUBMED									
REFERENCE									
AUTHORS									
TITLE									
JOURNAL									
MEDLINE									
PUBMED									
REFERENCE									
AUTHORS									
TITLE									
JOURNAL									
MEDLINE									
PUBMED									
REFERENCE									
AUTHORS									
TITLE									
JOURNAL									
MEDLINE									
PUBMED									
REFERENCE									
AUTHORS									
TITLE									
JOURNAL									
MEDLINE									
PUBMED									
REFERENCE									
AUTHORS									
TITLE									
JOURNAL									
MEDLINE									
PUBMED									
REFERENCE									
AUTHORS									
TITLE									
JOURNAL									
MEDLINE									
PUBMED									
REFERENCE									
AUTHORS									
TITLE									
JOURNAL									
MEDLINE									
PUBMED									
REFERENCE									
AUTHORS									
TITLE									
JOURNAL									
MEDLINE									
PUBMED									
REFERENCE									
AUTHORS									
TITLE									
JOURNAL									
MEDLINE									
PUBMED									
REFERENCE									
AUTHORS									
TITLE									
JOURNAL									
MEDLINE									
PUBMED									
REFERENCE									
AUTHORS									
TITLE									
JOURNAL									
MEDLINE									
PUBMED									
REFERENCE									
AUTHORS									
TITLE									
JOURNAL									
MEDLINE									

6 (bases 1 to 1699)
 Adachi,J., Aizawa,K., Akimura,T., Arakawa,T., Bono,H., Carninci,P., Fukuda,J.S., Furuno,M., Hanagaki,T., Hara,A., Hashizume,W., Hayashida,K., Hayatsu,N., Hiramoto,K., Hiraoka,T., Hirozane,T., Hori,F., Imotani,K., Ishii,Y., Itoh,M., Kagawa,I., Kasukawa,I., Katoh,H., Kawai,J., Kojima,Y., Kondo,S., Konno,H., Kouda,M., Koya,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Murata,M., Nakamura,M., Nishi,K., Nomura,K., Numazaki,R., Ohno,M., Ohsato,N., Okazaki,Y., Saito,R., Sano,H., Sakai,C., Sakai,K., Sakazume,N., Sano,H., Sasaki,D., Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Tagami,M., Tomaru,A., Takahashi,F., Takaku-Akahira,S., Takeda,Y., Tanaka,T., Tomaru,A., Toya,T., Yasunishi,A., Muramatsu,M. and Hayashizaki,Y.
 Direct Submission
 Submitted (16-APR-2002) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.go.jp, URL:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222, Fax:81-45-503-9216)
 cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.
 Please visit our web site for further details.
 URL:http://genome.gsc.riken.go.jp/
 URL:http://fantom.gsc.riken.go.jp/
 Location/Qualifiers
 1. .1699
 /organism="Mus musculus"
 /mol_type="mRNA"
 /strain="C57BL/6J"
 /db_xref="FANTOM_DB:7530415E19"
 /db_xref="MGI:2396949"
 /db_xref="taxon:10090"
 /clone="7530415E19"
 /sex="male"
 /tissue_type="eyeball"
 /clone_lib="RIKEN full-length enriched mouse cDNA library"
 /dev_stage="adult"
 184..1206
 /notes="unnamed protein product; PROSTASIN (SPTR|BAB82496, evidence: FASTY, 98.5%ID, 100%length, match=1020)
 putative"
 /codon_start=1
 /protein_id="BAC37362.1"
 /db_xref="GI:26347427"
 /translation="MALRVGILGQLEAVTILLGLLQSGIRADCEASCGAVIQPI
 ITGGGQWSPQWFWQVGIYDNGNCGSVLSNKKVYSAACHPREHRAVEYKLA
 DHQDYSYNDVTWVAQIIYHSSYREGSGQDIALKSPVFSRTYRIPCILPAAL
 ASFGNGHCTVYTGHWAVSVLQTPRPLQVLEPLIRETCSCLININAVEEERTHT
 QQDMLCAGYVGGDHQCSGSGSCPMEGIYWLGVISWGDACGAPNRPQVYVLT
 TVASWIIHHVAELQRPVVPQTESQPDGHLNCHHPFVSAAAPKLLRPVLFPLGLT
 GLLSLA"
 1671..1676
 /notes="putative"
 1699
 /notes="putative"
 polyA_signal
 polyA_site
 ORIGIN
 Query Match 17.7%; Score 194.6; DB 11; Length 1699;
 Best Local Similarity 53.5%; Pred. No. 1.9e-24;
 Matches 508; Conservative 0; Mismatches 419; Indels 22; Gaps 4
 QY 1 GGGCCCTGTCTGGGCATGGCCACAGAGGGGFCCTGGGCGCTGGGCAGCTGGGGCT 60
 Db 166 GGGCCCTGTCTAGGCATGGCCCTAAGGTTGGGCTGGGACTTGGGCAGCTGGAAGCT 225
 QY 61 GTGGCAATTCGACTCATCTACTCTTTACGGTGG--TGCCGTCGACCGCTAGGG 118
 Db 226 GTGACCAATCTGCTCTTCTCGGATGTCTCAGTCGCGAATCCGAGCTGAGGGACTGAA 285

JOURNAL
COMMENT

Unpublished (1999)

Contact: Robert Strausberg, Ph.D.

Email: c9abbs@mail.nih.gov

Tissue Procurement: Life Technologies, Inc.

cDNA Library Preparation: Life Technologies, Inc.

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Agencourt Bioscience Corporation

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

http://image.llnl.gov

Plate: LLAM12804 row: n column: 11

High quality sequence stop: 654.

Location/Qualifiers

FEATURES

source

1. .1013

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/clone="IMAGE:5759434"

/lab_host="DH10B"

/clone_lib="NIH MGC 116"

/note="Organ: pooled colon, kidney, stomach; Vector:

pCMV-SPORT6; Site 1: NotI; Site 2: EcoRV (destroyed); RNA

source anonymous pool of 3 colons, age 26 yo male, 49 yo

female, 71 yo male colon; 46 yo male kidney, and pool of 2

stomachs, 62 yo male and 70 yo female. Library is

oligo-dT primed and directionally cloned (EcoRV site is

destroyed upon cloning). Average insert size 1.4 kb,

insert size range 1-3 kb. Library is normalized and

enriched for full-length clones and was constructed by C.

Gruber (invitrogen). Research Genetics tracking code

023. Note: this is a NIH_MGC Library."

ORIGIN

Query Match 17.4%; Score 191.8; DB 12; Length 1013;

Best Local Similarity 55.5%; Pred. No. 5e-24;

Matches 420; Conservative 0; Mismatches 322; Indels 15; Gaps 2;

```
QY 1 GGGCCCTTGTCTCTGGGCGCATGGCCCGCAGAGGGGGTCTCTGGGGCTGGGCGAGCTGGGGGCT 60
DB 95 GGGCCCTTGTCTCTGGGCGCATGGCCCGCAGAGGGGGTCTCTGGGGCTGGGCGAGCTGGGGGCT 154
QY 61 GTGGCCAAATCTGACTCACTACTTACCTTTAGGGTGGTGGCGCTCCGACCGCTAGGGGC 120
DB 155 GTGGCCAAATCTGACTCACTACTTACCTTTAGGGTGGTGGCGCTCCGACCGCTAGGGGC 214
QY 121 CCCCGTACTCGGGCGCCCTGAGCCCTCGCCCGCATCTGGGGGGCTCAAAACGCGCAG 180
DB 215 GCTCCCTGGGTGTG-----GCCCGCCAGCAGCATCACAGTGGGGCGAGTGCAGTC 268
QY 181 CCGGGCACCTGGCTTGGCAAGTGAAGCTGCAACATGAGGTGGCCACATCTGGGGGGC 240
DB 269 GCGGGTCAAGTGGCGCTGGCAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAG 328
QY 241 TCCCTCATCGCCCTCTCTGGGCTCTCTCGGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 300
DB 329 TCTCTGCTGTGAGCAGTGGGTGCTGTGAGTGTCTGTGAGTGTCTGTGAGTGTCTGTGAGTGT 381
QY 301 TTGAGCGCCGCGCGAGTGGTCTGGGTGCTGAGTCTGGGGCGTGTGAGTCTGGGGCGGCGGCTG 360
DB 382 --GCACACACAGGAGCCTATGAGTCAAGCTGGGGGCGCCACACAGTACTTCTCTACTTCC 439
QY 361 GACGGCGCGCACACCGCGCGAGTGGCGCCCATCTGGTGGTGGCGGCCAATACAGCCAAATG 420
DB 440 GAGGACGCCAAGGTGAGCAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAG 499
QY 421 GAGTGGCGCGGACCTGGCCCTGTGGCTGTGGCTGTGGCTGTGGCTGTGGCTGTGGCTGTGGCTGT 480
DB 500 GGTCTCCAGGGCGCATTGCACTCTCTCACTACAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 559
QY 481 GTGTGGCTGTGTGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 540
DB 560 ATCGGCGCCCATCTGCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 619
```

```
QY 119 GCCCGCGTACTGCGGCGCCCTGAGCCCTCGGCGCCGATCGTGGGGGGCTCAAAACGCGC 178
DB 286 GCCTCTGTGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 345
QY 179 AGCGGGCAGCTGGCTTGGCAAGTGAAGCTGACCTGACCATGAGGTGGCGACATCTGGGGGG 238
DB 346 GGTGAG-----TGGCCCTTGGCAGTTCAGCATACCTACGATGGAACCATCTTGTGGCG 400
QY 239 GCTCCCTCATCGCCCGCCCTCTCGGTCTCTCTCGGTCTCTCTCTCTCTCTCTCTCTCTCTCT 298
DB 401 GGTGCTCGTGTCAAATAAATGGGTGTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGT 455
QY 299 GGTGGAGCCCGCGCGAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 358
DB 456 ----ACACAGCAGGGAAGCGTATGAGTGAAGCTGGGGGACACACAGCTAGACTCTCTACA 511
QY 359 TGGACGGCGGCACACCGCGCAGTGGCGCGCATCTGTGGTGGCGGCAATACAGCCAAAG 418
DB 512 GCAATGACATGTGTGTCAACAGTGGTCTGATCATCATCCACTCAAGCTACCGAAG 571
QY 419 TGGAGCTGGCGCGCAGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGG 478
DB 572 AGGCTCTCCAGGGGACATCGCTGTCTATCGCTTCAGCAGTCTGTCACTTCTCTCCGCT 631
QY 479 CCGTGTGGCTGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 538
DB 632 ACATCAGACCCATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 691
QY 539 GGGCCACCGGTGGGAGACGCTCCAGGAGCAGATCTCTGCTCTGCTGCTGCTGCTGCTGCTGCT 598
DB 592 CTGTACAGGATGGGTCTGTGTCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 751
QY 599 AGGAAGTGGAGCTAAGGTGTGTGGCGAGGCGACCTGTCAATGTCTTACAGCAGCGCG 658
DB 752 AGCAGCTCGAGTACCACTCATCATCGCGGGAAACCTGTAGCTGTGCTGTGCTGTGCTGTGCTG 811
QY 659 GTCCCTTCAACTCTCTCAGATATTGCCAGGATGCTGTGTGCTGTGCTGTGCTGTGCTGTGCT 718
DB 812 CGGTGCTTGAAGAACCGCACATATCCAGCAGGACATGCTGTGTGCTGTGCTGTGCTGTGCTGT 871
QY 719 GCGCAGGAGCACCTGCCAGGCTGACTCTGGGGGCGCCCTGCTGTGTGAGAAAGCGCGCC 778
DB 872 GAGGCAAGATGCTGCGCAGGCTGACTCTGGGGGCGCCCTGCTGTGTGCTGTGCTGTGCTGTGCT 931
QY 779 GCTGTTCCAGCAGGAATCAGCAGTGTGGTGTGGTGTGGTGTGGTGTGGTGTGGTGTGGTGTGG 838
DB 932 TCTGTTACTTGGCAGGCATTGTGAGTGGGGGTGATGCTGTGCTGTGCTGTGCTGTGCTGTGCTG 991
QY 839 GAGTGTTCACCTGCTGTGCTTACCTATGAGGATGATACGGAGCAGGTGATGGGTTCAG 896
DB 992 GAGTATACCTGTGACTTCTACTTACTTCTGCTGTGATCCACCATGTG-----GCAG 1045
QY 899 AGCTGGGCGCTGCTTCTCCCGAGCGCCCGAGAGCAGCAGTCAATG 947
DB 1046 AGCTCCAGCTCGAGTGTGTCCTCCAGACTCAGAGTCCCGCGGATG 1094
```

```
RESULT 7
BM923713
LOCUS BM923713 1013 bp mRNA linear EST 12-MAR-2002
DEFINITION AGENCOURT_6623973 NIH_MGC_116 Homo sapiens cDNA clone IMAGE:5759434
5' mRNA sequence.
ACCESSION BM923713
VERSION BM923713.1 GI:19374092
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1013)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
```

```

Qy 541 GCCACCGCTGGGAGAGCTCCAGGAGCGAGATCCTCTGCTCTCCCTGGTGGTCTACAG 600
Db 620 GTCACTGGCTGGGTCTGTGGCCCCCTCATGTAGCTCTCTGAGCCAGCCACTGCAG 679
Qy 601 GAAGTGAGCTAAGGCTGCTGGGCGAGCCACCTGTCTCAATCTCTCTACAGCCAGCCGGT 660
Db 680 CAACCTCGAGTGCCTCTGATCAGTCTGTGAGACGTGTAACTGCCTGTACAACATCGAGGCC 739
Qy 661 CCCTTCACCTCAGTCTCCAGATATGCGAGGATGCTGTGTGCTGTGGTACCCAGAGGGC 720
Db 740 AAGCTGAGGAGCGGCACATTTTTCACAGAGACATGGTGTGTGCTGTGGTATGTGGAGGG 799
Qy 721 CGCAGGACACCTCCAGAGGTGACTCTGGGGGGCCCC 757
Db 800 GGCAAGGAGCGCTGCCAGGCTGACTCTGGGGGGCCCC 836

RESULT 8
BB850795 433 bp mRNA linear EST 26-NOV-2001
LOCUS BB850795 RIKEN full-length enriched, adult inner ear Mus musculus
DEFINITION BB850795 cDNA clone F930109F10 5', mRNA sequence.
ACCESSION BB850795
VERSION BB850795.1 GI:17092249
KEYWORDS EST.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1. (bases 1 to 433)
Akimura,T., Arakawa,T., Hiramoto,K., Carninci,P., Furuno,M., Hanagaki,T.,
Hayatsu,N., Hayatsu,Y., Hirotsu,T., Hirotsu,T., Imotani,K.,
Ishii,Y., Ito,M., Kawai,J., Kojima,Y., Konno,H., Kouda,M.,
Matsuyama,T., Nakamura,M., Nishi,K., Nomura,K., Numasaki,R.,
Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K., Sakazume,N.,
Sasaki,D., Sato,K., Shibata,K., Shinagawa,A., Shiraki,T.,
Sogabe,Y., Suzuki,H., Tagawa,A., Takahashi,F., Takaku-Akahira,S.,
Tanaka,T., Tomaru,A., Toyota,T., Watahiki,A., Yasunishi,A.,
Muramatsu,M. and Hayashizaki,Y.
RIKEN Encyclopedia of Mouse Full-length cDNAs (Akimura,T., et al.
2001)
Unpublished (2001)
Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic
Sciences Center (GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@gsc.riken.go.jp,
URL: http://genome.gsc.riken.go.jp/
Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,
Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
prepare full-length cDNA libraries for rapid discovery of new
genes. Genome Res. 10 (10), 1617-1630 (2000)
wagi,K., Fujiwaka,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E.,
Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T.,
Matsuura,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A.
and Hayashizaki,Y.
RIKEN integrated sequence analysis (RISA) system--384-format
sequencing pipeline with 384 multicapillary sequencer. Genome Res.
10 (11), 1757-1771 (2000)
Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P.,
Sugahara,Y. and Hayashizaki,Y.
Computer-based methods for the mouse full-length cDNA
encyclopedia: real-time sequence clustering for construction of a
nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
Please visit our web site (http://genome.gsc.riken.go.jp) for
further details.
e mouse tissues.
FEATURES
Location/Qualifiers
1. .433

```

```

/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="F930109F10"
/tissue_type="inner ear"
/dev_stage="adult"
/clone_lib="RIKEN full-length enriched, adult inner ear"

ORIGIN

Query Match 17.4%; Score 191.2; DB 10; Length 433;
Best Local Similarity 86.7%; Pred. No. 4.9e-24;
Matches 222; Conservative 0; Mismatches 33; Indels 1; Gaps 1;

Qy 123 CCCGTACTGCGGGCGCCTGAGCCCTCGGCCCGCATCGTGGGGGGCTCAAAACGCGCAGCC 182
Db 178 CCTAGACTGGGGCGCCCTGAGCCCTCTTCCGCAATTGTGGGGGCTCGGACGCTCATCC 237
Qy 183 GGGCACCTGCGCTTGGCAAGTGAGCTGACCATGAGGAGGTGGCCACATCTGCGGGGGCTC 242
Db 238 TGGCACCTTGGCCATGGCAGGTGAGCTTGCATCAAGGTGGGGGCCACATCTGCGGGGGCTC 297
Qy 243 CCTCATCGCCCCCTCTGGGCTCTCTCGGCTGCTCACTGTTCATGACGATGGAGCTT 302
Db 298 CCTCATCGCCCCCTTCTGGGT-CTCTCGCTGCTCACTGTTCGAGCAATGAACCTT 356
Qy 303 GGAGCCCGCGCGAGTGGTGGTCTGCTGGGGCTGCACTCCCGAGACGGGGCCCTTGA 362
Db 357 GGAGCCCGCGGAGAAATGTGCTTCTGCTGGGGCTGCACTCCCGAGACGGGGCCCTTGA 416
Qy 363 CGGCGCGCACACCGCGC 378
Db 417 AGGAGCGCACATGGGC 432

RESULT 9
AK010640 1629 bp mRNA linear HTC 20-SEP-2003
LOCUS AK010640 Mus musculus ES cells cDNA, RIKEN full-length enriched library,
clone:2410039E18 product:PROTASIN, full insert sequence.
ACCESSION AK010640.1 GI:12846228
VERSION AK010640.1
KEYWORDS HTC; CAP trapper.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1. Carninci,P. and Hayashizaki,Y.
High-efficiency full-length cDNA cloning
Meth. Enzymol. 303, 19-44 (1999)
99279253
10349636
REFERENCE 2
AUTHORS Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,
Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
Normalization and subtraction of cap-trapper-selected cDNAs to
prepare full-length cDNA libraries for rapid discovery of new genes
Genome Res. 10 (10), 1617-1630 (2000)
20499374
11042159
REFERENCE 3
AUTHORS Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P.,
Konno,H., Akiyama,J., Nishi,K., Kitsuai,T., Tashiro,H., Itoh,M.,
Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A.,
Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K.,
Fujiwaka,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M.,
Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Kawai,J.,
Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
RIKEN integrated sequence analysis (RISA) system--384-format
sequencing pipeline with 384 multicapillary sequencer
Genome Res. 10 (11), 1757-1771 (2000)
20530913
JOURNAL
MEDLINE
TITLE
AUTHORS
JOURNAL
MEDLINE
TITLE
AUTHORS

```


AUTHORS

Akimura,T., Arakawa,T., Carninci,P., Furuno,M., Hanagaki,T., Hayatsu,N., Hiramoto,K., Hiraoka,T., Hirozane,T., Imotani,K., Ishii,Y., Ito,M., Kawai,J., Kojima,Y., Konno,H., Kouda,M., Matsuyama,T., Nakamura,M., Nishi,K., Nomura,K., Numasaki,R., Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K., Sakazume,N., Sasaki,D., Sato,K., Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H., Tagawa,A., Takahashi,F., Takaku-Akahira,S., Tanaka,T., Tomaru,A., Toya,T., Watahiki,A., Yasunishi,A., Muramatsu,M. and Hayashizaki,Y.

TITLE

RIKEN Encyclopedia of Mouse Full-length cDNAs (Akimura,T., et al. 2001)

JOURNAL

Unpublished (2001)

COMMENT

Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@gsc.riken.go.jp,
URL: http://genome.gsc.riken.go.jp/
Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. Genome Res. 10 (10), 1617-1630 (2000)
wagi,K., Fujiwaki,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsunaga,S., Kawai,O., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer. Genome Res. 10 (11), 1757-1771 (2000)
Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P., Sugahara,Y. and Hayashizaki,Y.
Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
Please visit our web site (http://genome.gsc.riken.go.jp) for further details.

FEATURES

source

Location/Qualifiers
1..429
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="F93010IE02"
/tissue_type="inner ear"
/dev_stage="adult"
/clone_lib="RIKEN full-length enriched, adult inner ear"

ORIGIN

Query Match 15.6%; Score 171.8; DB 10; Length 429;
Best Local Similarity 84.4%; Pred. No. 1.4e-20;
Matches 205; Conservative 0; Mismatches 37; Indels 1; Gaps 1;
QY 123 CCGTACTCGGGCGCCCTCGAGCCCTCGGCGCGATCGTGGGGGCTCAAGCGCGAGCC 182
DB 178 CTAGACTCGGGCGCCCTCGAGACCTCTCCCGCATGTGGGGGGGCTCGAGCGCTCATCC 237
QY 183 GGGCACCTGGCCCTTG-CAAGTGAGCGCTGCACATGGAGGTGGCCACATCTCGGGGGCT 241
DB 238 TGGCACTTGGCCATGGCCAGGTGAGCTGCATAAAGGTGGGGGGCAAAATCTCGGGGGCT 297
QY 242 CCCTCATCGCCCTCTCTGGGTCTCTCGCTGCTCAGCTTTTCATGAGCAATGGAGCT 301
DB 298 CCCTCATGACACCTTCTCTGGGTCTCTCGCTGCTCAGCTTTTCGTCGAGCAATGGAACCT 357
QY 302 TGGAGCCCGCGCGAGTGGTGGTACTGCTGGGGTGCACCTCCAGGACGCGGCCCTGG 361
DB 358 TGGAGCCCGCGAGCAATTGTGAGTTGTGCTGGGGTGCACCTCCAGGACGCGGCCCTGGA 417

QY 362 AGC 364

DB 418 AGG 420

RESULT 13

BX436299

LOCUS

DEFINITION BX436299 Homo sapiens THYMUS Homo sapiens cDNA clone CSOCAP001YG17
3-PRIME, mRNA sequence.

ACCESSION

VERSION BX436299

KEYWORDS

EST.

SOURCE

ORGANISM Homo sapiens (human)

REFERENCE

AUTHORS Li,W.B., Gruber,C., Jessee,J. and Polayes,D.

TITLE Full-length cDNA libraries and normalization

JOURNAL Unpublished (2001)

COMMENT Contact: Genoscope

Genoscope - Centre National de Sequencage

BP 191 91006 EVRY cedex - France

Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr

Library was constructed by Life Technologies, a division of

Invitrogen. This sequence belongs to sequence cluster 7995.f For

more information about this cluster, see

http://www.genoscope.cns.fr/

cgi-bin/cluster.cgi?seq=CSOCAP001AD09FM1&cluster=7995.f. Contact :

Feng liang Email : fliang@lifetech.com URL :

http://fulllength.invitrogen.com/ Invitrogen Corporation 1600

Faraday Avenue Genoscope sequence ID : CSOCAP001AD09FM1.

FEATURES

source

Location/Qualifiers

1..930

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/clone="CSOCAP001YG17"

/tissue_type="THYMUS"

/clone_lib="Homo sapiens THYMUS"

/note="Vector: pCMVSPORT 6; 1st strand cDNA was primed

with a NotI-oligo(dT) primer. Five prime end enriched,

double-strand cDNA was digested with Not I and cloned into

the Not I and EcoRV sites of the pCMVSPORT 6 vector.

Library was not normalized."

ORIGIN

Query Match 15.6%; Score 171.4; DB 13; Length 930;
Best Local Similarity 99.4%; Pred. No. 2.1e-20;
Matches 172; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 773 GCGCGCGCTGGTCCAGCGAGGAATCACCAGCTTTGGCTGTGCGAGGAAACC 832
DB 68 GCGCGCGCTGGTCCAGCGAGGAATCACCAGCTTTGGCTGTGCGAGGAAACC 127
QY 833 GCCCTGGAGTTTTCACCTGCTGCTGCTACCTATGAGGCATGATACGGGAGCAGGTGATGG 892
DB 128 GCCCTGGAGTTTTCACCTGCTGCTGCTACCTATGAGGCATGATACGGGAGCAGGTGATGG 187
QY 893 GTTCAGAGCCTGGGCGCTGCTTTCCACCCAGCCCGAGAACCCAGTCAGAT 945
DB 188 GTTCAGAGCCTGGGCGCTGCTTTCCACCCAGCCCGAGAACCCAGTCAGAT 240

RESULT 14

BBS93614

LOCUS

DEFINITION BBS93614 RIKEN full-length enriched, 4 days neonate male adipose
Mus musculus cDNA clone B430104M11 5', mRNA sequence.

ACCESSION

BBS93614

VERSION

BBS93614.1

KEYWORDS

EST.

SOURCE
ORGANISM
Mus musculus (house mouse)

REFERENCE
AUTHORS
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 249)

AUTHORS
Aizawa, K., Akahira, S., Akimura, T., Arai, A., Arakawa, T., Carninci, P., Hanagaki, T., Hayatsu, N., Hiraoka, T., Hirozane, T., Hodojima, Y., Imotani, K., Ishii, Y., Itoh, M., Izawa, M., Kawai, J., Kojima, Y., Konno, H., Kusakabe, M., Matsuyama, T., Miyazaki, A., Nakamura, M., Nishi, K., Nomura, K., Numazaki, R., Okazaki, A., Okido, T., Owa, C., Sakai, C., Sakai, K., Sasaki, D., Sato, K., Shibata, K., Shibata, Y., Shingawa, A., Shiraki, T., Sogabe, Y., Suzuki, H., Tagawa, A., Takahashi, F., Tanaka, T., Toya, T., Matakiki, A., Yamamura, T., Yasunishi, A., Yoshida, K., Yoshiki, A., Muramatsu, M. and Hayashizaki, Y.

TITLE
JOURNAL
COMMENT
RIKEN Mouse ESTs (Aizawa, K. et al. 2000)
Unpublished (2000)
Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Teurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-resgsc.riken.go.jp,
URL: http://genome.gsc.riken.go.jp/
Carninci, P., Nishiyama, Y., Westover, A., Itoh, M., Nagaoka, S., Sasaki, N., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
Thermotabilization and thermoactivation of thermolabile enzymes by trehalose and its application for the synthesis of full length cDNA. Proc. Natl. Acad. Sci. U.S.A. 95 (2), 520-524 (1998)
Itoh, M., Kitsuina, T., Akiyama, J., Shibata, K., Izawa, M., Kawai, J., Tomaru, Y., Carninci, P., Shibata, Y., Ozawa, Y., Muramatsu, M., Okazaki, Y. and Hayashizaki, Y.
Automated filtration-based high-throughput plasmid preparation system. Genome Res. 9 (5), 463-470 (1999)
Carninci, P. and Hayashizaki, Y.
High-efficiency full-length cDNA cloning. Methods Enzymol. 303, 19-44 (1999)
Please visit our web site (<http://genome.rtc.riken.go.jp>) for further details.

FEATURES
source
1..249
/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="taxon:10090"
/clone="B430104M11"
/sex="male"
/tissue_type="adipose"
/dev_stage="4 days neonate"
/lab_host="DH10B"
/clone_lib="RIKEN full-length enriched, 4 days neonate male adipose"
/note="Site 1: SalI; Site 2: BamHI; cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer [5' GAGAGAGAGATTCGAGTTAATAATTAATCCGCCCCCCCCCC sequence [5' GAGAGAGATTCGAGTTAATAATTAATCCGCCCCCCCCCC 3']. cDNA was cleaved with XhoI and BamHI. Vector: a Lambda modified pBluescript KS(+) after bulk excision from Lambda FLIC I."

ORIGIN
Query Match
15.4%; Score 169.6; DB 10; Length 249;

Best Local Similarity 80.2%; Pred. No. 2.8e-20;
Matches 199; Conservative 0; Mismatches 49; Indels 0; Gaps 0;

QY 381 AGTGGCGGCATCGTGGTGCCTGCGGCAACTACAGCAAGTGGAGCTGGGCGCCGACTGGC 440
Db 2 AGTGGCGGCATCGTGGTGCCTGCGGCAACTACAGCAAGTGGAGCTGGGCGCCGACTGGC 61
QY 441 CTGTGCTGGCGCTGGCGCTGCGGCAACTACAGCAAGTGGAGCTGGGCGCCGACTGGC 500
Db 62 CTGTGCTGGCGCTGGCGCTGCGGCAACTACAGCAAGTGGAGCTGGGCGCCGACTGGC 121
QY 501 CGCGCGCTCACACCGCTGCTGTCACCGCACTGGCGCTGGCGCTGGCGCTGGCGCTGGC 560
Db 122 TGTGCTGCTGCAACTCTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 191
QY 561 CCAGGAGGAGGAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 620
Db 182 CCAGGAGGAGGAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 241
QY 621 GGGCGGAGG 628
Db 242 GGGGAGG 249

RESULT 15
AZ252377
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

AZ252377 430 bp DNA linear GSS 15-JUN-2000
RPCI-23-466N15.TVB RPCI-23 Mus musculus genomic clone
RPCI-23-466N15, genomic survey sequence.
AZ252377
AZ252377.1 GI:8565590
GSS.
Mus musculus (house mouse)
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
Zhao, S., Nierman, W., Feldblyum, T., Malek, J., Shatsman, S., Akinret, B., Levins, M., McGann, S., Tsegaye, G., Geer, K., Krol, M., de Jong, P. and Fraser, C.M.
Mouse BAC End Sequences from Library RPCI-23
Unpublished (1999)
Other GSSs: RPCI-23-466N15.TJ RPCI-23-466N15.TV RPCI-23-466N15.TVB
Contact: Shaying Zhao
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: szhao@tigr.org
Clones are derived from the mouse BAC library RPCI-23. For BAC library availability, please contact Pieter de Jong (pieter@jeong.med.buffalo.edu). Clones may be purchased from BACPAC Resources (<http://bacpac.med.buffalo.edu/orderingframe.htm>) or from Resea ch Genetics (info@resgen.com). BAC end page: http://www.tigr.org/tdb/bac_ends/mouse/bac_end_intro.html
Plate: 466 row: N column: 15
Seq primer: SP6
Class: BAC ends.

FEATURES
Location/Qualifiers
1..430
/organism="Mus musculus"
/mol_type="genomic DNA"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="RPCI-23-466N15"
/sex="Female"
/lab_host="DH10B"
/clone_lib="RPCI-23"
/note="Organ: Kidney/Brain; Vector: pBACe3 6; Site 1: EcoRI; Site 2: EcoRI; Female C57BL/6J mouse kidney and/or brain genomic DNA was isolated and partially digested with a combination of EcoRI and EcoRI Methylase. Size

selected DNA was cloned into the pBACe3.6 vector at the
EcoRI sites. The ligation products were transformed into
DH10B electrocompetent cells (BRL Life Technologies). "

ORIGIN

Query Match 15.3%; Score 168.6; DB 28; Length 430;
Best Local Similarity 88.4%; Pred. No. 5.1e-20;
Matches 183; Conservative 0; Mismatches 24; Indels 0; Gaps 0;
QY 737 AGGTTGACTCTGGGGGCCCCCTGGTCTGTGAGGAGGCGCGCTGGTTCCAGGACAGGAA 796
Db |||||||||||||||| |||||||||||||||| |||||||||||||||| ||||||||||||||||
40 AGGTTGACTCTGGGGGACCGCTGGTCTGTGAGGAGGCGCGCTGGTTCCAGGACAGGAA 99
QY 797 TCACCAAGCTTTGGTTTGGCTGTGGACGAGAGAAACCGCCCTGGAGTTCACCTGTGTGG 856
Db |||||||||||||||| |||||||||||||||| |||||||||||||||| ||||||||||||||||
100 TCACCAAGCTTTGGCTTTGGCTGTGGACGAGAGAAACCGCCCTGGAGTTCACCTGTGTGG 159
QY 857 CTACCTATGAGGCATGGATACGGGACGAGTGTATGGGTTTCAGAGCCTGGGCTTTCCTTC 916
Db |||||||||||||||| |||||||||||||||| |||||||||||||||| ||||||||||||||||
160 CTCCCTATGAGTCGTGGATTTCGGGACACGTGTATGGGTTTCAGAACCTGGGCTTTCCTTC 219
QY 917 CCACCCAGCCCCAGAGACCCAGTCAG 943
Db |||||||||||||||| |||||||||||||||| |||||||||||||||| ||||||||||||||||
220 CCAGCCAGCTCCAGAGCCACAGTCAG 246

Search completed: February 25, 2004, 11:17:42
Job time : 2791 secs

Result No.	Score	Query %		Length	DB	ID	Description
		Match					
1	224	20.3	1613	4	US-09-387-375-1		Sequence 1, Appli
2	216.8	19.7	1130	4	US-09-387-375-8		Sequence 8, Appli
3	183.2	16.6	1142	4	US-09-386-642-8		Sequence 8, Appli
4	183.2	16.6	1169	4	US-09-386-642-7		Sequence 7, Appli
5	165.2	15.0	1110	4	US-09-386-653A-1		Sequence 1, Appli
6	164.2	14.9	980	4	US-09-023-942A-30		Sequence 30, Appli
7	162	14.7	1212	4	US-09-620-312D-431		Sequence 431, App
8	161.6	14.7	1101	4	US-09-023-942A-5		Sequence 5, Appli
9	161.6	14.7	1100	4	US-09-907-794A-256		Sequence 256, App
10	161.6	14.7	1100	4	US-09-905-125A-256		Sequence 256, App
11	161.6	14.7	1100	4	US-09-902-775A-256		Sequence 256, App
12	160	14.5	1081	3	US-09-008-271A-15		Sequence 15, Appl
13	157.6	14.3	959	4	US-09-023-942A-25		Sequence 25, Appl
14	154.8	14.0	1094	4	US-09-023-942A-3		Sequence 3, Appli
15	153.8	14.0	1130	4	US-09-386-653A-8		Sequence 8, Appli
16	145.8	13.2	1103	4	US-09-386-642-59		Sequence 59, Appl
17	141.6	12.8	3147	2	US-09-027-337-1		Sequence 1, Appli
18	141.6	12.8	3147	4	US-09-644-600-1		Sequence 1, Appli
C 19	141.6	12.8	3147	4	US-09-644-600-18		Sequence 18, Appl
20	141.6	12.8	3147	4	US-09-654-600A-1		Sequence 1, Appli
C 21	141.6	12.8	3147	4	US-09-654-600A-18		Sequence 18, Appl
22	140.6	12.8	2152	4	US-09-023-655-157		Sequence 157, App
23	139.8	12.7	796	4	US-09-280-116-107		Sequence 107, App
24	139.8	12.7	1081	2	US-09-016-366A-22		Sequence 22, Appl
25	139.8	12.7	1081	2	US-08-978-404B-17		Sequence 17, Appl
26	139.8	12.7	1137	2	US-09-016-366A-18		Sequence 18, Appl
27	139.8	12.7	1137	2	US-08-978-404B-13		Sequence 13, Appl


```

1  RESULT 8
2  US-09-023-942A-5
3  ; Sequence 5, Application US/09023942A
4  ; Patent No. 6479274
5  ; GENERAL INFORMATION:
6  ; APPLICANT: (US only) ANTALIS Toni Marie and HOOPER John David
7  ; TITLE OF INVENTION: NOVEL MOLECULES
8  ; NUMBER OF SEQUENCES: 30
9  ; CORRESPONDENCE ADDRESS:
10 ; ADDRESSEE: SCULLY, SCOTT, MURPHY & PRESSER
11 ; STREET: 400 GARDEN CITY PLAZA
12 ; CITY: GARDEN CITY
13 ; STATE: NEW YORK
14 ; COUNTRY: USA
15 ; ZIP: 11530
16 ; COMPUTER READABLE FORM:
17 ; MEDIUM TYPE: Floppy disk
18 ; COMPUTER: IBM PC compatible
19 ; OPERATING SYSTEM: PC-DOS/MS-DOS
20 ; SOFTWARE: PatentIn Release #1.0, Version #1.25
21 ; CURRENT APPLICATION DATA:
22 ; APPLICATION NUMBER: US/09/023,942A
23 ; FILING DATE: 13-FEB-1998
24 ; CLASSIFICATION: 435
25 ; PRIOR APPLICATION DATA:
26 ; APPLICATION NUMBER: P05101/97
27 ; FILING DATE: 13-FEB-1997
28 ; PRIOR APPLICATION DATA:
29 ; APPLICATION NUMBER: PP0422/97
30 ; FILING DATE: 18-NOV-1997
31 ; PRIOR APPLICATION DATA:

```

Query Match	14.7%;	Score 162;	DB 4;	Length 1212;
Best Local Similarity	53.6%;	Pred. No. 9.4e-28;		
Matches	414;	Conservative 0;	Mismatches 340;	Indels 18; Gaps 3;
QY	112	GCTAGGGGCCCCGTA	CTCGGGCGGCGCTT	GAGCCCTCGGCCCGCATCGTGGGGGGCTCA 171
Db	192	GCCAAGGCAGACAC	GCGCTTGTTCGCCCC	CAGATGCTGAACCGATGGTGGCGGGCAG 251
QY	172	AACGCGAGCGCGGC	ACCTCGCCCTTGC	CAAGTGAAGCTGCACCATGGAGTGGCCACATC 231
Db	252	GACACGACGAGGG	CGAGTGGCCCTTGC	CAAGTCAAGATCCAGCGGCACACGGAAGCCATTC 311
QY	232	TGCGGGGGTCCCT	CAATCGCCCGCCCT	CTCTCGGGTCTCTTCGCGTGTCTCACTCTTTT---CATG 288
Db	312	TGCGGGGGCAGC	CTCATCGCGAGAGT	GGGTCTCTGACGCTCGGCACTGCTCTCCCGAAC 371

APPLICATION NUMBER: International PCT Application
FILING DATE: 13-FEB-1998
ATTORNEY/AGENT INFORMATION:
NAME: DIGILIO, FRANK S
REGISTRATION NUMBER: 31,346
REFERENCE/DOCKET NUMBER: 11168
TELEPHONE: (516) 742 4343
TELEFAX: (516) 742 4366
TELECOMMUNICATION INFORMATION:
TELEX: 230 901 SANS UR
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 1100 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA
FEATURE:
NAME/KEY: CDS
LOCATION: 17..961
US-09-023-942A-5

Query Match 14.7%; Score 161.6; DB 4; Length 1100;
Best Local Similarity 52.9%; Pred. No. 1.1e-27;
Matches 406; Conservative 0; Mismatches 344; Indels 18; Gaps 2;
QY 130 TGGGGGCGCTGAGCCCTCGGCGCGCATCTGGGGGCTCAACGCGCGAGCGGCGACC 189
DB 113 TGGGGCGAGCGGTCTATCATCTGCGCGCATCTGGGGTGGAGGAGCGCGCACTGGGCGGT 172
QY 190 TGGCCCTTGGCAAGTGAAGCTGACCATGAGAGTGGCCACATCTCGGGGGCTCCCTCATC 249
DB 173 TGGCCCTTGGCAAGTGAAGCTGACCATGAGAGTGGCCACATCTCGGGGGCTCCCTCATC 232
QY 250 GCGCCCTCTGGTCTCTCCCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 309
DB 233 AGCCACCGCTGGGCACTCAAGCGCGCGCACTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 232
QY 310 GCGCGCGAGTGGTCT 360
DB 293 CCCTCCGGTGGTGTCT 352
QY 361 GAGCGCGCGACACCGCGCGAGTGGCGCGCATCTGGGTGGCGCGCACTACAGCGCAAGTG 420
DB 353 CAGCGCTACTACCGCTTACTTCT 412
QY 421 GAGCTGGGCGCGACCTGGCCCT 480
DB 413 AATTACCTTATGACATCT 472
QY 481 GTGTGGCT 540
DB 473 ATCCAGCCCATCT 532
QY 541 GCGACCGGCTGGGAGAGCT 600
DB 533 GTGACTGGCTGGGAGTACATCAAGAGGATGAGGACCTGCCATCTCTCCACACCCCTCCAG 592
QY 601 GAAGTGAGCTAAGGCTGTGGCGGAGGCGCACTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 660
DB 593 GAAGTTGAGTCTGCGCATCAACCACTCTATGTGCAACCACTCTCTCTCTCTCTCTCTCTCT 652
QY 661 CCCTTAACT 720
DB 653 TTCCGCAAGGAC-----ATCTTTGAGAGCATGGTTTGTGCTGGCAATGCCCAAGGC 703
QY 721 CGCAGGACACCTGCCAGGTGACTCTGGGGGCGCCCTGCTCTGTGAGGAAGGGCGGCGCG 780
DB 704 GGGAGAGATGCT 763
QY 781 TGGTCCAGGAGGAATCACCACCTTTGGTTGGTGTGGAGCGAGAAACGGCCCTGGA 840
DB 764 TGGTATCAGATTGAGTCTGGGAGTGGGCTGTGGTGGGCGGCAATCGGCGCGCGGT 823

QY 841 GTTTTCACCTGCTGTGGCTTACCTATGAGCGCATGGATACGGGAGCGAGTG 888
DB 824 GTCTACCAATATCAGCCACCACTTTTGGTGGATCCAGAAAGCTGATG 871
RESULT 9
US-09-907-794A-256
Sequence 256, Application US/09907794A
Patent No. 6635468
GENERAL INFORMATION:
APPLICANT: Genentech, Inc.
APPLICANT: Ashkenazi, Avi
APPLICANT: Botstein, David
APPLICANT: Desnoyers, Luc
APPLICANT: Eaton, Dan L.
APPLICANT: Ferrara, Napoleone
APPLICANT: Filvaroff, Ellen
APPLICANT: Fong, Sherman
APPLICANT: Gao, Wei-Qiang
APPLICANT: Gerber, Hanspeter
APPLICANT: Gerritsen, Mary E.
APPLICANT: Goddard, A.
APPLICANT: Godowski, Paul J.
APPLICANT: Grimaldi, Christopher J.
APPLICANT: Gurney, Austin L.
APPLICANT: Hillan, Kenneth J.
APPLICANT: Kljavin, Ivar J.
APPLICANT: Mather, Jennie P.
APPLICANT: Pan, James
APPLICANT: Paoni, Nicholas F.
APPLICANT: Roy, Margaret Ann
APPLICANT: Stewart, Timothy A.
APPLICANT: Tumas, Daniel
APPLICANT: Williams, P. Mickey
APPLICANT: Wood, William, I.
TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
TITLE OF INVENTION: Acids Encoding the Same
FILE REFERENCE: 10466-14
CURRENT APPLICATION NUMBER: US/09/907,794A
CURRENT FILING DATE: 2001-07-17
PRIOR APPLICATION NUMBER: PCT/US00/04414
PRIOR FILING DATE: 2000-02-22
PRIOR APPLICATION NUMBER: US 60/143,048
PRIOR FILING DATE: 1999-07-07
PRIOR APPLICATION NUMBER: US 60/145,698
PRIOR FILING DATE: 1999-07-26
PRIOR APPLICATION NUMBER: US 60/146,222
PRIOR FILING DATE: 1999-07-28
PRIOR APPLICATION NUMBER: PCT/US99/20594
PRIOR FILING DATE: 1999-09-08
PRIOR APPLICATION NUMBER: PCT/US99/20944
PRIOR FILING DATE: 1999-09-13
PRIOR APPLICATION NUMBER: PCT/US99/21090
PRIOR FILING DATE: 1999-09-15
PRIOR APPLICATION NUMBER: PCT/US99/21547
PRIOR FILING DATE: 1999-09-15
PRIOR APPLICATION NUMBER: PCT/US99/23089
PRIOR FILING DATE: 1999-10-05
PRIOR APPLICATION NUMBER: PCT/US99/28214
PRIOR FILING DATE: 1999-11-29
PRIOR APPLICATION NUMBER: PCT/US99/28313
PRIOR FILING DATE: 1999-11-30
PRIOR APPLICATION NUMBER: PCT/US99/28564
PRIOR FILING DATE: 1999-12-02
PRIOR APPLICATION NUMBER: PCT/US99/28565
PRIOR FILING DATE: 1999-12-02
PRIOR APPLICATION NUMBER: PCT/US99/30095
PRIOR FILING DATE: 1999-12-16
PRIOR APPLICATION NUMBER: PCT/US99/30911
PRIOR FILING DATE: 1999-12-20
PRIOR APPLICATION NUMBER: PCT/US99/30999
PRIOR FILING DATE: 1999-12-20

Thu Feb 26 13:50:23 2004

```

; PRIOR APPLICATION NUMBER: PCT/US00/00219
; PRIOR FILING DATE: 2000-01-05
; NUMBER OF SEQ ID NOS: 423
; SEQ ID NO 256
; LENGTH: 1100
; TYPE: DNA
; ORGANISM: Homo Sapien
; US-09-907-794A-256

Query Match      14.7%; Score 161.6; DB 4; Length 1100;
Best Local Similarity 52.9%; Pred. No. 1.1e-27;
Matches 406; Conservative 0; Mismatches 344; Indels 18; Gaps 2;

QY 130 TGGCGGCGCCTGAGCCTGCGCCCGCCTGCGGGGGCTCAACGCGCAGCGGGSCACC 189
DB 116 TGGCGGCGCAGGCTCATCAGTCGCGCATCTGGTGGGAGAGACCGCAATCTCGGGCT 175

QY 190 TGGCCCTTGGCACTGAGCTGACCATGAGGAGTGGCCACATCTGCGGGGGCTCCCTCATC 249
DB 176 TGGCCCTTGGCGGGAGCCTTGCCTGTGGGATTCACAGTATGCGGAGTGAGCCTGTCTC 235

QY 250 GCCCCCTCTCGGTCTCTCCGTCTCTCACTGTTTATGACGAATGGGAGCTTGGAGCCC 309
DB 236 AGCCACCGCTGGGCACTCAACGCGGCGCACTGCTTTGAAACCTATAGTACCTTAGTGAT 295

QY 310 GCGGCGGAGTGGTGGTACTGCTGGGGCTGC-----ACTCCAGGAGCGGCCCTG 360
DB 296 CCTCCGGGTGGATGTTGTCAGTCTGGGCGAGCTGACTTCCATGCCATCTCTCTGAGCGCTG 355

QY 361 GAGCGGCGCACACCGCGCAGTGGCGGCCATCTGTGTGGCGGCCAACTACAGCCAAAGTG 420
DB 356 CAGGCTTACTACACCCCTTACTTGTATCGAATATCTATCTGAGCCCTGCTACCTGGGG 415

QY 421 GAGCTGGCGCGCAGCTGGCCCTGCTGGCTGTCGCTGACCCGCGAGCCTGGGCGCCGCC 480
DB 416 AATTACCCCTATGACATTTGCCCTTGGTGAAGTGTCTGCACCTGTCACTTAAACAC 475

QY 481 GTGTGGCTGTGCTGCTGCGCCGCGCTCACACGCTTCTGTGACGCGCAGCGCTGCTGG 540
DB 476 ATCCAGCCCATCTGTCTCCAGGCTCCACATTTGAGTTTGAGAACCGGACAGACTGTGG 535

QY 541 GCCACCGCTGGGAGAGCTGCAGGAGCAGATCTCTGCTCTCTCCCTGGGTGCTACAG 600
DB 536 GTGACTGCTGGGGTACATCAAGAGATGAGGACCTGCCATCTCCCAACAGCTCCAG 595

QY 601 GAAGTGGAGCTAAGGCTGCTGGGAGGCGCCACTGTCTCAATGCTCTAGCAGCCCGCT 660
DB 596 GAAGTTCAGGTGCGCCATCAATAAACAACTCTATGTGCAACCACTCTTCTCAAGTACAGT 655

QY 661 CCCTTCAACTCACTCTCCAGATATTCAGGAGTGTGTGCTGGCTACCCAGAGGGC 720
DB 656 TTCCGCAAGGAC-----ATCTTTGGAGACATGGTTTGTGCTGGCAACGCCCAAGGC 706

QY 721 CGCAGGACACCTGCGCAGGCTCACTCTGGGGGGCCCTGCTGTGTGAGAGCGGCGCGC 780
DB 707 GGAAGGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 766

QY 781 TGGTTTCCAGGCGGAATCAACAGCTTGGTGTGGCTGTGGAGGAGAACCGCTTGA 840
DB 767 TGGTATCAGATTGGAGTCTGAGCTGGGAGTGGGCTGTGTGCTGGGCGCAATCGGCCGCT 826

QY 841 GTTTTCACTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 888
DB 827 GTCTACCAATATCAGCCACCACTTTGAGTGGATTCAGAGTGTGATG 874
```

```

RESULT 10
US-09-905-125A-256
; Sequence 256, Application US/0905125A
; Patent No. 6664376
; GENERAL INFORMATION:
; APPLICANT: Genentech, Inc.
; APPLICANT: Ashkenazi, Avi
```

```

; APPLICANT: Botstein, David
; APPLICANT: Desnovers, Luc
; APPLICANT: Eaton, Dan L.
; APPLICANT: Ferrara, Napoleone
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Fong, Sherman
; APPLICANT: Gerber, Hanspeter
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, A.
; APPLICANT: Godowski, Paul J.
; APPLICANT: Grimaldi, Christopher J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Hillan, Kenneth, J.
; APPLICANT: Kljavin, Ivar J.
; APPLICANT: Mather, Jennie P.
; APPLICANT: Pan, James
; APPLICANT: Paoni, Nicholas F.
; APPLICANT: Roy, Margaret Ann
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Williams, P. Mickey
; APPLICANT: Wood, William, I.
; TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
; FILE REFERENCE: 10466-14
; CURRENT APPLICATION NUMBER: US/09/905,125A
; CURRENT FILING DATE: 2001-07-12
; PRIOR APPLICATION NUMBER: PCT/US00/04414
; PRIOR FILING DATE: 2000-02-22
; PRIOR APPLICATION NUMBER: US 60/143,048
; PRIOR FILING DATE: 1999-07-07
; PRIOR APPLICATION NUMBER: US 60/145,698
; PRIOR FILING DATE: 1999-07-26
; PRIOR APPLICATION NUMBER: US 60/146,222
; PRIOR FILING DATE: 1999-07-28
; PRIOR APPLICATION NUMBER: PCT/US99/20594
; PRIOR FILING DATE: 1999-09-08
; PRIOR APPLICATION NUMBER: PCT/US99/20944
; PRIOR FILING DATE: 1999-09-13
; PRIOR APPLICATION NUMBER: PCT/US99/21090
; PRIOR FILING DATE: 1999-09-15
; PRIOR APPLICATION NUMBER: PCT/US99/21547
; PRIOR FILING DATE: 1999-09-15
; PRIOR APPLICATION NUMBER: PCT/US99/23089
; PRIOR FILING DATE: 1999-10-05
; PRIOR APPLICATION NUMBER: PCT/US99/28214
; PRIOR FILING DATE: 1999-11-29
; PRIOR APPLICATION NUMBER: PCT/US99/28313
; PRIOR FILING DATE: 1999-11-30
; PRIOR APPLICATION NUMBER: PCT/US99/28564
; PRIOR FILING DATE: 1999-12-02
; PRIOR APPLICATION NUMBER: PCT/US99/28565
; PRIOR FILING DATE: 1999-12-02
; PRIOR APPLICATION NUMBER: PCT/US99/30095
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: PCT/US99/30911
; PRIOR FILING DATE: 1999-12-20
; PRIOR APPLICATION NUMBER: PCT/US99/30999
; PRIOR FILING DATE: 1999-12-20
; PRIOR APPLICATION NUMBER: PCT/US00/00219
; PRIOR FILING DATE: 2000-01-05
; NUMBER OF SEQ ID NOS: 423
; SEQ ID NO 256
; LENGTH: 1100
; TYPE: DNA
; ORGANISM: Homo Sapien
; US-09-905-125A-256

Query Match      14.7%; Score 161.6; DB 4; Length 1100;
Best Local Similarity 52.9%; Pred. No. 1.1e-27;
Matches 406; Conservative 0; Mismatches 344; Indels 18; Gaps 2;
```


Db 296 CCCTCCGGGTGAGTGTCCAGTTTGGCCAGCTGACTTCCATGCCATCTTCTGGAGCCTG 355
QY 361 GAGGGCGGACACCCGCGCAGTGGCGGCATCGTGGTCCGCCCACTACAGCCAAAGTG 420
Db 356 CAGGCTACTACACCCGTTACTTCGTATCGAATATATCTGAGCCCTCGCTACCTGGGG 415
QY 421 GAGCTGGGCGCCGACCTGGCCCTGCTGCGCTGGCCCTCAACCCGAGCCTGGGCGCCGCG 480
Db 416 AATTACACCTATGACATGCTTGGTGAAGCTCTGACCTCTCACCTACACTAAACAC 475
QY 481 GTGTGGCTGTCTGGCTGCCCCCGGCTCACACGCTTGTGTGACGCGACCGCTGCTGG 540
Db 476 ATCCAGCCCATCTGTCTCCAGGCTCCACATTTGAGTTTGAGAAACCGGACAGACTGCTGG 535
QY 541 GCCACCGCTGGGAGACGCTCCAGGAGGCGATCCTGCTGCTCCCTGGGTGCTACAG 600
Db 536 GTGACTGGCTGGGGGTACATCAAGAGGATGAGCACTGCCATCTCCCAACACCTCCAG 595
QY 601 GAAGTGAAGCTAAGGCTGTGGCGAGGCCACCTGTCAATGTCTCTACAGCCAGCCCGGT 660
Db 596 GAAGTTCAAGTCCCATCAATAAACAATCTATGTGCAACCACTCTTCTCAAGTACAGT 655
QY 661 CCCTTCAACCTCACTCCAGATATTCGCGAGGATGCTGTGTGCTGCTACCCAGAGGC 720
Db 656 TTCGCGAAGGAC-----ATCTTTGGAGACATGGTTTGTGTGCGCAACGCCCAAGGC 706
QY 721 CGCAGGACACCTGCCAGGTGACTCTGGGGGCCCTGTGTGTGAGGAAGCGCGCGC 780
Db 707 GGAAGATGACCTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 766
QY 781 TGGTTCCAGCAGGAATACACAGCTTTGGGTTTGGCTGTGGAGCGAGAAACCGCCCTGGA 840
Db 767 TGTATCAGATTGAGTGTGAGCTGGGAGTGGCTGTGGTGGTGGTGGTGGTGGTGGTGGT 826
QY 841 GTTTTCACTGCTGTGCTACCTATGAGGCAATGATAGGAGGAGGAGTG 888
Db 827 GTCTACCAATATCAGCCACCACTTTGAGTGGATCCAGAAGCTGATG 874

RESULT 12

US-09-008-271A-15
; Sequence 15, Application US/09008271A
; Patent No. 6203979
; GENERAL INFORMATION:
; APPLICANT: Bandman, Olga
; Hillman, Jennifer L.
; Yue, Henry
; Guegler, Karl J.
; Corley, Neil C.
; Tang, Tom Y.
; Shah, Purvi
; TITLE OF INVENTION: HUMAN PROTEASE MOLECULES
; NUMBER OF SEQUENCES: 24
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Dr.
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/008,271A
; FILING DATE: 16-Jan-1998
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: <Unknown>
; FILING DATE: <Unknown>
; ATTORNEY/AGENT INFORMATION:

NAME: Mohan-Peterson, Sheela
REGISTRATION NUMBER: 41,201
REFERENCE/DOCKET NUMBER: PF-0458 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 650-855-0555
TELEFAX: 650-845-4166
INFORMATION FOR SEQ ID NO: 15:
SEQUENCE CHARACTERISTICS:
LENGTH: 1081 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY: PROSTUT03
CLONE: 789927
SEQUENCE DESCRIPTION: SEQ ID NO: 15 :
US-09-008-271A-15
Query Match 14.5%; Score 160; DB 3; Length 1081;
Best Local Similarity 52.7%; Pred. No. 2,6e-27;
Matches 405; Conservative 0; Mismatches 345; Indels 18; Gaps 2;
QY 130 TGGGGCGCCCTGAGCCCTCGGCGCCGATCGTGGGGGCTCAACCGCAGCGCGGACC 189
Db 136 TGGGGCGCCGAGGCTCATCAGTCGCGCATCGTGGGTGGAGAGCGCGAATCGGGCGGT 195
QY 190 TGGCCCTTGGCAAGTGAAGCTGACCTGCACCATGAGGTGGCCACATCTGGGGGGCTCCCTCATC 249
Db 196 TGGCGGTGGCAGGAGAGCTTGGCCCTGTGGGATCCACGTATGCGAGTGAAGCTGCTC 255
QY 250 GCGCCCTCTCGGTCTCTCTCGCTGCTCACTCTTTATGCAAGATGGGACGTTTGGAGCCC 309
Db 256 AGCCACCGCTGGGCACTCAACGCGCGCACTGCTTTGAAACCTATAGTGACCTTAGTGAT 315
QY 310 GCGGCGAGTGTGGTACTGTGGGCGTGC-----ACTCCAGACGCGGCGCGCTG 360
Db 316 CCCTCGGGTGGATGGTCCAGTTTGGCCAGCTGACTTCCATGCCAATCTTCTGGAGCCCTG 375
QY 361 GACGGCGGCACACCGCGCAGTGGCGGCATCTGTGGTCCGCGCAACTACACAGCCAAAGTG 420
Db 376 GAGGCTACTACACCGCTTACTTGTATCGAATATCTATCTGAGCCCTCGCTACCTGGGG 435
QY 421 GAGCTGGGCGCGACCTGGCCCTGTGCGCTGCGCTCAACCGCCAGCTTGGGCGCCCGCC 480
Db 436 AATTCACCCCTATGACATTCCTTGGTGAAGCTGTCTGCACTGTCACTTACCTAAACAC 495
QY 481 GTGTGCGCTGTGCTGCTGCGCGCCCTCACACCGCTTGTGTCAGCGCACCGCTGCTGG 540
Db 496 ATCCAGCCCATCTGTCTCCAGGCGCTCCACATTTGAGTTTGGAAACCGGACAGACTGCTGG 555
QY 541 GCCACCGCTGGGAGACGCTCCAGGAGGAGATCTCTGCTCTCCCTTGGGTGCTTACAG 600
Db 556 GTGACTGGCTGGGGGTACATCAAGAGGATGAGGCACTGCCATCTCCCAACACCTCCAG 615
QY 601 GAAGTGAAGCTAAGGCTGTGGCGGAGGCCAAGCTGTCAATGTCTTACAGCAGCGCCGCT 660
Db 616 GAAGTTCAAGTGGCCCATCAATAAACAACCTATATGTGCAACCACTCTTCTTCAAGTACAGT 675
QY 661 CCCTTCAACCTCACTCTCCAGATATTCAGGAGGATGCTGTGTGCTGCTTACCCAGAGGC 720
Db 676 TTCGCAAGGAC-----ATCTTTGAGACATATGGTTTGTGGGCAATGCCCAAGGC 726
QY 721 CGCAGGACACCTGGCAGGCTGACTCTGGGGGGCGCCCTGCTGTGAGGAAGCGCGCGCC 780
Db 727 GGAAGGATGCTTGTCTGGTGAAGTCAAGTGAAGCTTGGCTGTAAACAGGATGAGCTG 786
QY 781 TGGTTCCAGCGAGGAATACAGCTTTGGGTTTGGCTGTGAGCGGAGAAACCGCCCTGGA 840
Db 787 TGTATCAGATTGAGTGTGAGCTGGGAGTGGGCTGTGTGCGGCCCAATCGGCGCGCT 846
QY 841 GTTTTCACTGCTGTGCTACCTATGAGGCAATGATACGGGAGCAGTG 888
Db 847 GTCTACCAATATCAGCCACCACTTTGAGTGGATCCAGAAGCTGATG 894

RESULT 13

US-09-023-942A-25

; Sequence 25, Application US/09023942A

; Patent No. 6479274

; GENERAL INFORMATION:

; APPLICANT: (US only) ANTALIS Toni Marie and HOOPER John David

; TITLE OF INVENTION: NOVEL MOLECULES

; NUMBER OF SEQUENCES: 30

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: SCULLY, SCOTT, MURPHY & PRESSER

; STREET: 400 GARDEN CITY PLAZA

; CITY: GARDEN CITY

; STATE: NEW YORK

; COUNTRY: USA

; ZIP: 11530

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/09/023,942A

; FILING DATE: 13-FEB-1998

; CLASSIFICATION: 435

; PRIOR APPLICATION NUMBER: P05101/97

; FILING DATE: 13-FEB-1997

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: P0422/97

; FILING DATE: 18-NOV-1997

; APPLICATION DATA:

; APPLICATION NUMBER: International PCT Application

; FILING DATE: 13-FEB-1998

; ATTORNEY/AGENT INFORMATION:

; NAME: DIGIGLIO, FRANK S

; REGISTRATION NUMBER: 31,346

; REFERENCE/DOCKET NUMBER: 11168

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (516) 742 4343

; TELEFAX: (516) 742 4366

; TELEX: 230 901 SANS UR

; INFORMATION FOR SEQ ID NO: 25:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 959 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; MOLECULE TYPE: DNA

; FEATURE:

; NAME/KEY: CDS

; LOCATION: 2..856

US-09-023-942A-25

Query Match

Best Local Similarity 14.3%; Score 157.6; DB 4; Length 959;

Matches 377; Conservative 0; Mismatches 334; Indels 9; Gaps 1;

QY	129	CTGGCGGCGCCCTGAGCCCTCGGCGCGCATCGTGGGGGCTCAGACGCGCGCGGCAC	188
Db	19	CTGGCGGTACAGACATCCCTTCCTGATAGTGGGTGGCGATGATGCTGAGCTTGGCGG	78
QY	189	CTGGCGTTGGCAAGTGAAGCTGACCTGACCATGAGAGTGGCCACATCTCGGGGGGTCCCTCAT	248
Db	79	CTGGCGGTGGCAAGGAGGCTCGTGTATGGGCAACCATCTATGTGGCGCAACCTTCT	138
QY	249	CGCCCGCTCTGGGCTCTCTCGCTGCTCATCTGTTTCATGAGAAATGGGACGTTGGAGCC	308
Db	139	CAACCGCGCTGGGTGCTTACAGCTGCCCACTGCTTCCAAAAGGATAACGATCCCTTTGA	198
QY	309	CGCGCGGAGTGGTGGTACTCTGCGGTGTCACCTCCCGAGACGCGGCCCTGGAGGGCG	368

Db	199	CTGGACAGTCCAGTTTGGTGAGCTGACTTCAGGCCATCTCTCTGGAACCTACAGGCCTA	258
QY	369	GCACACCGCGCAGTGGCCGCCCATGTGGTGGCGGCCAACTACAGCAAGCTGGAGCTGGG	428
Db	259	TTCCACCGGTTACCAATAGAGATATTTTCTGAGCCCCCAAGTACTCGGAGCAGTATCC	318
QY	429	CGCGACCTGGCCCTGTGGCGCTGGGCTCACCGCCGCGCAGCTGGGCCGCCCGCTGGCC	488
Db	319	CAATGACATAGCCCTGCTGAAGCTGTCTCCAGTCACTTCACTACATAACTTCAATCCAGCC	378
QY	489	TGCTGCTGCCCCCGCCCTCACACCGCTTGTGTCAGCGCACCGCCTGCTGGGCGCACCGG	548
Db	379	CATCTGCTCTCTGAAGTCCACGCTACAAAGTTTGAGAACCGAACTGACTGCTGGGTGACCGG	438
QY	549	CTGGGGAGACGCTCCAGGAGCAGATCTCTGCTCTCTCCCTGGTGTCTACAGGAAGTGA	608
Db	439	CTGGGGGCTATTGGAGAGATGAGAGTCTGCCATCTCCCAACTCTCCAGGAGTGA	498
QY	609	GCTAAGGCTCTGGCGGAGGCCACCTGTCAATGTCTACAGCCAGCCCGCTCCCTTCAA	668
Db	499	GGTAGCTATTATCAACAAACAGCATGTCTAACCATATGTACAAAAGCCAG-----A	549
QY	669	CTTCACTCTCCAGATATTGCCAGGATGCTGTGCTGTGCTGCTACCCAGAGGCCCGCAGGA	728
Db	550	CTTCCGCAACAACATCTGGGAGACATGGTTGGCTGGCACTCTGAAGGTGCAAGGA	609
QY	729	CACCTGCAGGTCACCTCTGGGGGGCCCTGGTCTGTGAGGAAGCGCGCTGGTTCCA	788
Db	610	TGCTGCTTTGGTGACTTCGGGAGGACCTTGGCTGGCCAGCATACGGTGTGTATCA	669
QY	789	GGCAGGAATCACAGCTTTGGGTTTGGCTGTGACGCGAGAAACCGCCCTGGAGTTTCAAC	848
Db	670	GTTTGGAGTTGTGAGCTGGGGAATAGGCTGTGCTGGCCCAATCGCCCTGAGTCTATAC	729

RESULT 14

US-09-023-942A-3

; Sequence 3, Application US/09023942A

; Patent No. 6479274

; GENERAL INFORMATION:

; APPLICANT: (US only) ANTALIS Toni Marie and HOOPER John David

; TITLE OF INVENTION: NOVEL MOLECULES

; NUMBER OF SEQUENCES: 30

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: SCULLY, SCOTT, MURPHY & PRESSER

; STREET: 400 GARDEN CITY PLAZA

; CITY: GARDEN CITY

; STATE: NEW YORK

; COUNTRY: USA

; ZIP: 11530

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/09/023,942A

; FILING DATE: 13-FEB-1998

; CLASSIFICATION: 435

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: P05101/97

; FILING DATE: 13-FEB-1997

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: P0422/97

; FILING DATE: 18-NOV-1997

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: International PCT Application

; FILING DATE: 13-FEB-1998

; ATTORNEY/AGENT INFORMATION:

; NAME: DIGIGLIO, FRANK S

; REGISTRATION NUMBER: 31,346

; REFERENCE/DOCKET NUMBER: 11168

; TELECOMMUNICATION INFORMATION:

US-09-386-653A-8
; Sequence 8, Application US/09386653A
; Patent No. 6458564
; GENERAL INFORMATION:
; APPLICANT: Andrade-Gordon, Patricia
; APPLICANT: Darrow, Andrew
; APPLICANT: Qi, Jian-shen
; TITLE OF INVENTION: DNA encoding the novel human serine
; FILE REFERENCE: ORT-1032
; CURRENT APPLICATION NUMBER: US/09/386.653A
; CURRENT FILING DATE: 1999-08-31
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 8
; LENGTH: 1130
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Fusion gene of
; OTHER INFORMATION: Protease T in a zymogen activation vector
US-09-386-653A-8

Query Match 14.0%; Score 153.8; DB 4; Length 1130;
Best Local Similarity 53.9%; Pred. No. 6.6e-26;
Matches 392; Conservative 0; Mismatches 317; Indels 18; Gaps 3;

157 ATCTGGGGGGCTCAACGCGCAGCGGGGACCTGGCCCTTGGCAAGTGGAGCTGCACCAT 216
166 ATCTGGGGGGCTATGCTCTAGAGGAGGCGAGTGGCCCTTGGCAAGTGGAGCTGCACCAT 225
217 GGAGTGGCGCAATCTGGGGGGCTCCCTATATGCGCCCTTCTTGGGTCTCTCCCTGCT 276
226 AACGGAAGCCATTCTGGGGGGGAGCCCTCATCGCGAGCAGTGGGTCTCTACGCGCTGG 285
277 CACTGTTT---CATGACGAATGGGACCTTGGAGCCGCGGCGAGTGGTTCGTAATGCTG 333
286 CACTGCTTCCGCAACACCTCTGAGACGCTCCCTGTATACAGGTCTCTGCGGGCAAGGCG 345
334 GCGGTGCACTCCAGGACGGGCGCTGACGCGCGGCGCACACCGCGCGAGTGGCCCGCATC 393
346 CTAGTGGAGCGCGGACACACGCTATGATGCCCGGTGAGGAGTGGAGAGCAACCC 405
394 GTGGTGGCGGCAACTACAGCCAGTGGAGTGGCGGCGGAGCTGGCCCTTGTGGCGCTG 453
406 CTGTACCAAGGCA-----CGGCGCTCCAGCGCTGACGTGGCCCTGTGGAGCTG 453
454 GCCTACCGCGCAGCTGGGCGCGCGCTGTGGCTGTCTGCTGCGCCCGCTCAGC 513
454 GAGGCACAGTGGCTTCAACATTTACATCTCCCGGTGTGCTGCTGACCCCTCGGTG 513
514 CGCTTCGTGCAAGGCAACCGCTGTGGGCGCACCGCTTGGGGAGAGCTCCAGAGGAGAT 573
514 ATCTTGGAGCGGGCATGAACTGTGGGTCACTGTGGGCGAGCCCGCTGAGGAGAGAC 573
574 CTTCTGCTCTCCCTGGGTGCTACAGGAAGTGGAGCTTAAGGTGCTGGCGAGGCCACC 633
574 CTTCTGCGCGAACCAGCGGATCTCTGAGAACTCTGCTGTGGTGTGCTGATCATGACAC 633
634 TGTCAATGCTCTACAGCGAGCCCGCTCCCTTCAACCTCACTCTCCAGATATTT---GCCA 690
634 TGAACCTGCTCTACAGCAAGACACCGAGTTTGGTGTACCAACCCCAACCAATCAAGAT 693
691 GGGATGCTGTGTGCTGGCTTACCCAGAGGCGCGAGGACACCTGCGAGGGTGAATCTGGG 750
694 GACATGCTGTGCGCGCTTTCGAGGAGGSCAAGAGGATGCTGCAAGGGGAGCTCGGCG 753
751 GGGCCCTGCTGTGTGAGGAGCGCGCTGCTGCTTCCAGGAGGAATCACAGCTTTGGG 810
754 GGGCCCTGCTGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 813
811 TTTGCTGTGGAGGAGAAACCGCTGAGTTTTCACCTGTGTGCTGCTGCTGCTGCTGCTGCT 870

TELEPHONE: (516) 742 4343
TELEFAX: (516) 742 4366
TELEX: 230 901 SANS UR
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1094 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 17..955
US-09-023-942A-3

Query Match 14.0%; Score 154.8; DB 4; Length 1094;
Best Local Similarity 52.2%; Pred. No. 3.9e-26;
Matches 398; Conservative 0; Mismatches 352; Indels 12; Gaps 2;

130 TCGCGCGCGCTGAGCCCTCGCGCCGCTCGTGGGGGGCTCAACGCGCAGCGCGGCGCAC 189
113 TCGCGCGAGGGTCACTACGTCGCGCATCGTGGTGGAGAGAGCGCCGAATCTGGCGGT 172
190 TGGCTTGGCAAGTGAAGCTGACCATGAGAGTGGCCACATCTGGGGGGCTCCCTCATC 249
173 TGGCCGTGGCAGGGAGCGTGGCCCTGTGGGATTCACAGTATGCGGAGTGAAGCTGCTC 232
250 GCCCCTCTGAGTCTCTCCGCTGCTCACTCTTCAATGACGAAGTGGAGCGCTGGAGCC 309
233 AGCCACCGCTGGGCACTACGGCGGGCGCACTGCTTGAACCTGACCTTAGTATCCCTCC 292
310 G---GCGCGAGTGGTGGTACTGCTGGGCGTGCATCTCCAGGACGGGCGCTGGAGCGC 366
293 GGGTGAATGCTCCAGTTTGGCCAGCTGATCTCCATGCCATCTCTTGGAGCGCTGAGGCC 352
367 GGGCAACCGCGAGTGGCGCGCAATGCTGCTGGTGGCGGCGCAATACAGCCAAATGAGCTG 426
353 TACTACACCGCTTACTTCTGATCGAATATCTATCTGAGCGCTCGTACCTGGGGAATTC 412
427 GCGCGCACTGGCGCTGCTGCGCTGCGCTGCGCTGCGCTGCGCTGCGCTGCGCTGCG 486
413 CCGTATGACATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 472
487 CTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 546
473 CCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 532
547 GGGTGGGAGAGCTCCAGGAGGAGATCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 606
533 GGGTGGGAGATCAACAGAGGATGAGGCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 592
607 GAGCTAAGGCTGCTGGCGGAGGCGACCTGCTCAATGCTCTTACAGCGAGCGCGCTGCT 666
593 CAGTGGCGCATCAACAACTCTATGTCGACCACTCTCTCTCAAGTACAGTTTCCGC 652
667 AACCTCACTCTCCAGATATGCGAGGAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 726
653 AAGGAC-----ATCTTGGAGACATGCTTGTGCTGGCAATGCCCAAGGCGGGAAG 703
727 GACACTGCGAGGAGTCACTTGGGGGGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 786
704 GATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 763
787 CAGGAGGAGTCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 846
764 CAGATTGGAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 823
847 ACTGCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 888
824 ACCAATATCAGCCACCACTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 865

Db 814 GAGGCTGTGCCCCGACAGAACCGCCAGGTGTCTACATCGTGTACCGGCCCAAC 873
QY 871 TGGATAC 877
| | | | |
Db 874 TGGATCC 880

Search completed: February 25, 2004, 11:19:47
Job time : 118 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 25, 2004, 10:31:35 ; Search time 452 Seconds

(without alignments)
8537.195 Million cell updates/sec

Title: US-10-037-417-45

Perfect score: 1102

Sequence: 1 gggccctgtctggccat.....gggtctgtatggggcctcc 1102

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 2308694 seqs, 1750822206 residues

Total number of hits satisfying chosen parameters: 4617368

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications NA.*

- 1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq.*
- 2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq.*
- 3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq.*
- 4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq.*
- 5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq.*
- 6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq.*
- 7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq.*
- 8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq.*
- 9: /cgn2_6/ptodata/1/pubpna/US09A_PUBCOMB.seq.*
- 10: /cgn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq.*
- 11: /cgn2_6/ptodata/1/pubpna/US09C_PUBCOMB.seq.*
- 12: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq.*
- 13: /cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq.*
- 14: /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq.*
- 15: /cgn2_6/ptodata/1/pubpna/US10C_PUBCOMB.seq.*
- 16: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq.*
- 17: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq.*
- 18: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	815.6	74.0	2457	9	US-09-888-615-52
2	232.4	21.1	3382	14	US-10-101-510-447
3	230.8	20.9	1733	14	US-10-176-847-85
4	225.6	20.5	1020	15	US-10-051-874-25
5	224	20.3	1613	13	US-10-041-400A-1
6	224	20.3	1613	13	US-10-041-264A-1
7	224	20.3	1613	13	US-10-042-091A-1
8	221.2	20.1	1834	9	US-09-948-094-1
9	221.2	20.1	1834	9	US-09-860-107-2214
10	221.2	20.1	1834	9	US-09-967-768A-141
11	221.2	20.1	1834	14	US-10-097-340-261
12	220.6	20.0	1668	9	US-09-925-301-208
13	216.8	19.7	1130	13	US-10-041-400A-8
14	216.8	19.7	1130	13	US-10-041-264A-8
15	216.8	19.7	1130	13	US-10-042-091A-8

16	135.2	17.7	1797	14	US-10-109-616-1	Sequence 1, Appli
17	185.6	16.8	882	12	US-10-042-865-33	Sequence 33, Appl
18	184	16.7	882	12	US-10-042-865-34	Sequence 34, Appl
19	180.8	16.4	768	14	US-10-221-097-10	Sequence 10, Appl
20	179	16.2	1161	12	US-10-042-865-31	Sequence 31, Appl
21	176.2	16.0	1327	9	US-09-978-295A-170	Sequence 170, App
22	176.2	16.0	1327	9	US-09-978-697-170	Sequence 170, App
23	176.2	16.0	1327	9	US-09-978-192A-170	Sequence 170, App
24	176.2	16.0	1327	9	US-09-999-832A-170	Sequence 170, App
25	176.2	16.0	1327	10	US-09-978-189-170	Sequence 170, App
26	176.2	16.0	1327	10	US-09-978-608A-170	Sequence 170, App
27	176.2	16.0	1327	10	US-09-978-585A-170	Sequence 170, App
28	176.2	16.0	1327	10	US-09-978-191A-170	Sequence 170, App
29	176.2	16.0	1327	10	US-09-978-403A-170	Sequence 170, App
30	176.2	16.0	1327	10	US-09-978-564A-170	Sequence 170, App
31	176.2	16.0	1327	10	US-09-999-833A-170	Sequence 170, App
32	176.2	16.0	1327	10	US-09-981-915A-170	Sequence 170, App
33	176.2	16.0	1327	10	US-09-978-824-170	Sequence 170, App
34	176.2	16.0	1327	10	US-09-918-585A-170	Sequence 170, App
35	176.2	16.0	1327	10	US-09-978-423A-170	Sequence 170, App
36	176.2	16.0	1327	10	US-09-978-193A-170	Sequence 170, App
37	176.2	16.0	1327	10	US-09-999-830A-170	Sequence 170, App
38	176.2	16.0	1327	10	US-09-978-757A-170	Sequence 170, App
39	176.2	16.0	1327	10	US-09-978-187B-170	Sequence 170, App
40	176.2	16.0	1327	10	US-09-978-643A-170	Sequence 170, App
41	176.2	16.0	1327	10	US-09-978-375A-170	Sequence 170, App
42	176.2	16.0	1327	10	US-09-978-298A-170	Sequence 170, App
43	176.2	16.0	1327	10	US-09-978-188A-170	Sequence 170, App
44	176.2	16.0	1327	10	US-09-978-681A-170	Sequence 170, App
45	176.2	16.0	1327	10	US-09-978-194A-170	Sequence 170, App

ALIGNMENTS

RESULT 1

US-09-888-615-52
; Sequence 52, Application US/09888615
; Patent No. US2002064856A1
; GENERAL INFORMATION:
; APPLICANT: FLOWMAN, GREGORY
; APPLICANT: WHYTE, DAVID
; APPLICANT: CHARENEPEL, SEAN
; APPLICANT: CHARYDCZAK, GLEN
; APPLICANT: MANNING, GERARD
; APPLICANT: SUDARSANAM, SUCHA
; TITLE OF INVENTION: NOVEL PROTEASES
; FILE REFERENCE: 038602/1214
; CURRENT APPLICATION NUMBER: US/09/888,615
; CURRENT FILING DATE: 2001-06-26
; PRIOR APPLICATION NUMBER: 60/214,047
; PRIOR FILING DATE: 2000-06-26
; NUMBER OF SEQ ID NOS: 150
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 52
; LENGTH: 2457
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-888-615-52

Query Match 74.0%; Score 815.6; DB 9; Length 2457;
Best Local Similarity 99.5%; Pred. No. 1.3e-204;
Matches 818; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY	124	CGGTACTGCGGGCGCCCTGAGCCCTCGGCCCGCATCGTGGGGGGCTCAAAACGCGACGCG 183
Db	106	CTGGATGCGGGCGCCCTGAGCCCTCGGCCCGCATCGTGGGGGGCTCAAAACGCGACGCG 165
QY	184	GGCACCCTGCGCTTGGCAAGTGAGCCCTGCAACATGAGAGTGGCCACATCTCGCGGGGCTCC 243
Db	166	GGCACCTGCGCTTGGCAAGTGAGCCCTGCAACATGAGAGTGGCCACATCTCTCGCGGGGCTCC 225
QY	244	CTCATCGCCCCCTCTGGGTCTCTCCGCTGCTCATTCTTCATGACGAATGGGACGTTG 303

Db 226 CTATGCGCCCTCTCTGGGTCCTCTCGCTGCTCACTGTTTCATGACGAATGGAGCGCTG 285
Qy 304 GAGCCGCGCGCGAGTGGTGGTACCTGCTGGGCGTGCACCTCCAGAGACGGGCGCCCTGGAC 363
Db 286 GAGCCGCGCGCGAGTGGTGGTACCTGCTGGGCGTGCACCTCCAGAGACGGGCGCCCTGGAC 345
Qy 364 GCGCGCACACCGCGCGAGTGGGCGCATGCTGGTGCAGGCGGCGGCGGCGGCGGCGGCGG 423
Db 346 GCGCGCACACCGCGCGAGTGGGCGCATGCTGGTGCAGGCGGCGGCGGCGGCGGCGGCGG 405
Qy 424 CTGGGCGCGGACCTGGGCGCTGCTGGGCGTGCACCTCCAGGCGGCGGCGGCGGCGGCGG 483
Db 406 CTGGGCGCGGACCTGGGCGCTGCTGGGCGTGCACCTCCAGGCGGCGGCGGCGGCGGCGG 465
Qy 484 TGGGCTGTGCTGCTGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 543
Db 466 TGGGCTGTGCTGCTGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 525
Qy 544 ACCGGCTGGGAGAGCTCCAGAGGCGAGATCTCTGCTGCTCTCCCTGGGTCGTACAGGAA 603
Db 526 ACCGGCTGGGAGAGCTCCAGAGGCGAGATCTCTGCTGCTCTCCCTGGGTCGTACAGGAA 585
Qy 604 GTGAGCTTAAGGCTGTGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 663
Db 586 GTGAGCTTAAGGCTGTGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 645
Qy 664 TTCAACCTCACTCTCCAGATATGCGAGGAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 723
Db 646 TTCAACCTCACTCTCCAGATATGCGAGGAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 705
Qy 724 AGGACACCTGCGAGGAGT 783
Db 706 AGGACACCTGCGAGGAGT 765
Qy 784 TTCCAGGCGAGGAATCACAGCTTTGGGTTTGGCTGTGGACCGGAGAAACCGGCTTGGAGTT 843
Db 766 TTCCAGGCGAGGAATCACAGCTTTGGGTTTGGCTGTGGACCGGAGAAACCGGCTTGGAGTT 825
Qy 844 TTCACTGCTGTGGCTTACCTATGAGCATGAGATACGGAGCAGGATGATGGTTTCAGAGCT 903
Db 826 TTCACTGCTGTGGCTTACCTATGAGCATGAGATACGGAGCAGGATGATGGTTTCAGAGCT 885
Qy 904 GGGGCTGCTTTCCACCGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 945
Db 886 GGGGCTGCTTTCCACCGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 927

RESULT 2
US-10-101-510-447
; Sequence 447, Application US/10101510
; Publication No. US20030148295A1
; GENERAL INFORMATION:
; APPLICANT: WAN, JACKSON
; APPLICANT: WANG, YIXIN
; TITLE OF INVENTION: EXPRESSION PROFILES AND METHODS OF USE
; FILE REFERENCE: 15117.0012
; CURRENT APPLICATION NUMBER: US/10/101,510
; CURRENT FILING DATE: 2002-03-20
; PRIOR APPLICATION NUMBER: 60/276,947
; PRIOR FILING DATE: 2001-03-20
; NUMBER OF SEQ ID NOS: 805
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 447
; LENGTH: 3382
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-101-510-447

Query Match 21.1%; Score 232.4; DB 14; Length 3382;
Best Local Similarity 55.0%; Pred. No. 2.4e-51;
Matches 508; Conservative 0; Mismatches 401; Indels 15; Gaps 2;

Qy 1 GGGCCCTCTCTCTGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 60
Db 212 GGGCCCTCTCTCTGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 271
Qy 61 GTGGCCAAATTCATGACATCTACATCTTACGGGTTGGTGGCGGCGGCGGCGGCGGCGGCGG 120
Db 272 GTGGCCAAATTCATGACATCTTACATCTTACGGGTTGGTGGCGGCGGCGGCGGCGGCGGCGG 331
Qy 121 CCCCCCTACTCTGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 180
Db 332 GCTCCCTCTGGGTTGT-----GCCCGCCAGACGCGATCACAGTGGCAGCATGCGATC 385
Qy 181 CCGGGCACCTGGGCTTTGGCAAGTGAAGCTGCACCAATGAGAGTGGCGCAATCTCGCGGCGG 240
Db 386 GCGGTCAGTGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 445
Qy 241 TCCTCATGCGCCCTCTCTGGGTCCTCTCGGCTGCTCACTGTTTCATGACGAATGGGACG 300
Db 446 TCTCTCTGTGTGAGCAGTGGGTCGTGAGCTGCTCAGCTGCTCAGCTGCTTCCCGCAGCGA----- 498
Qy 301 TTGGAGCCCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 360
Db 499 --GCACCAAGAGAGCTATGAGGTCAAGCTGGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 556
Qy 361 GACGGCGCGCACACCGCGCGAGTGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 420
Db 557 GAGGCGCGCAAGGTGAGCAGCCCTGAGGACATCATCCCCCAGCCAGCTACCTCCAGGAG 616
Qy 421 GAGTGGGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 480
Db 617 GGTCCCGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 676
Qy 481 GTGTGGGCTGTGCTGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 540
Db 677 ATCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 736
Qy 541 GCGACCGGCTGGGAGAGCTGCGAGGAGAGATCTCTGCGCTCTCCCTGCGGTCGCTACAG 600
Db 737 GTCACTGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 796
Qy 601 GAGTGGAGCTAAGGCTGCTGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 660
Db 797 CACTCGAGGTGCTCTGATGATGCTGAGACGCTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTG 856
Qy 661 CCTTTCAACCTCACTCTCCAGATATGCGAGGAGTGTGTGTGTGTGTGTGTGTGTGTGTGT 720
Db 857 AAGCCTGAGGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 916
Qy 721 CGCAGGAGACCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 780
Db 917 GCGAAGGAGCGGCTGCGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCTC 976
Qy 781 TGGTTCCAGGCGGAGATCACAGCTTTGGGTTTGGGCTGTGAGCGGAGAAACCGGCGGCGG 840
Db 977 TGGTACCTGACGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1036
Qy 841 GTTTTCACTGCTGTGGCTACTATGAGGCGATGAGATACGGGAGCAGGATGATGGGTTTCAGAG 900
Db 1037 GTGTACACTCTGGGCTCCAGCTATGCTCTCTGATGATGATGATGATGATGATGATGATGATG 1096
Qy 901 CCGGCGGCTGCTTCCCGCCAG 924
Db 1097 CCGGCTGTGTGGTGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1120

RESULT 3
US-10-176-847-85
; Sequence 85, Application US/10176847
; Publication No. US20030068636A1
; GENERAL INFORMATION:
; APPLICANT: Veiby, Petter Ole
; TITLE OF INVENTION: COMPOSITIONS, KITS, AND METHODS FOR
; IDENTIFICATION, ASSESSMENT, PREVENTION, AND THERAPY OF BREAST

; TITLE OF INVENTION: AND OVARIAN CANCER
; FILE REFERENCE: MEI-039
; CURRENT APPLICATION NUMBER: US/10/176,847
; CURRENT FILING DATE: 2002-06-21
; NUMBER OF SEQ ID NOS: 112
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 85
; LENGTH: 1733
; TYPE: DNA
; ORGANISM: Homo sapiens
; US-10-176-847-85

Query Match 20.9%; Score 230.8; DB 14; Length 1733;
Best Local Similarity 54.9%; Pred. No. 5.9e-51;
Matches 507; Conservative 0; Mismatches 402; Indels 15; Gaps 2;

Qy	1	GGGCGCTTGTCTGGCCCATGGCCAGAGGGGTCTCTGGGGCTTGGGCAGCTGGGGCT	60
Db	81	GGGCGCTTGTCTGGCCCATGGCCAGAGGGGTCTCTGGGGCTTGGGCAGCTGGGGCT	140
Qy	61	GTGGCCATTCTGACTCATCTCACTTTACGGGTGTGGTCCGTCGGACCCGCTAGGGGC	120
Db	141	GTGGCCATTCTGCTCTATCTTGGATTACTCGGTTCAGGACAGAGCGGAGGGGCAGAA	200
Qy	121	CCCCGCTACTGGGGCGCCCTGAGCCCTCGGCCGCAATCGTGGGGGCTCAAAACGCGCAG	180
Db	201	GCTCCCTCGGTGTG-----GCCCGCCCAAGCAGCATCACAGGTGGCAGCAGTGCGATC	254
Qy	181	CGGGCAGCTGGCTTGGCAAGTGAAGCTGACCATGACCATGAGGTGGCCACATCTGGGGGGC	240
Db	255	GCGGTCAGTGGCCCTGGCAGGTGACGATCATGATGAGGCGTCCATGTGTGTGGTGGC	314
Qy	241	TCCCTCATGCCCCCTCTCGGTCTCTCCGCTGCTCACTGTTTCATGACGAATGGGACG	300
Db	315	TCTCTGCTGTGACGAGTGGTGTGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG	367
Qy	301	TTGAGCCCGGGCGAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG	360
Db	368	--GCACACAGAGAGCCTATGAGTCAAGCTGAGTGGGGGCCACAGCTAGACTCTCTACTC	425
Qy	361	GACGGGGGCAACCGCGGAGTGGCGGCAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG	420
Db	426	GAGGACGCGCAAGGTGAGCAGCCTGAGGAGCATCATCCCCCAGCCAGCTTACCTCCAGGAG	485
Qy	421	GAGCTGGGGCGGAGCTGCGCTGCTGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG	480
Db	486	GGCTCCGAGGGGCAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG	545
Qy	481	GTGTGGCTGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG	540
Db	546	ATCCGGCCCATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG	605
Qy	541	GCCACGGCTGGGAGAGCTTCAGGAGGAGATCTCTGCTGCTGCTGCTGCTGCTGCTGCTG	600
Db	606	GTCACTGGCTGGGGTATGTGGCCCGCTCAGTGAGCTCTGAGCGCCCAAGCCATCTGAG	665
Qy	601	GAAGTGGAGCTAAGGTGCTGGCGGAGGCACTGTCAATGTCTTACAGCCAGCCCGCT	660
Db	666	CAACTGAGGTGCTCTGATCAGTGTGAGACGTGTAACTGCTGTACCAATCGAGGCC	725
Qy	661	CCCTTCAACCTCACTTCCAGATATTGCCAGGATGCTGTGTGCTGGCTTACCCAGAGGC	720
Db	726	AAGCCTGAGAGCGGACCTTTGTCAGAGGACATGCTGTGTGCTGCTGCTGCTGCTGCTG	785
Qy	721	CGCAGGACACCTGCCAGGTGACTCTGGGGGGCCCTGCTGTGTGAGGAGGCGCGCCG	780
Db	786	GGCAAGAGCGCTTCCAGGGTGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG	845
Qy	781	TGCTTCCAGCAGGAATCACCACTTGGTGTGGTGTGGTGTGGTGTGGTGTGGTGTGGTGT	840
Db	846	TGCTTCCAGCAGGAATGTGAGCTGGGGAGATGCTGTGTGGGGCCCGCAGAGGCTTGT	905
Qy	841	GTTTTCACTGCTGTGGCTACCTATGAGGCATGATGATGATGATGATGATGATGATGATG	900

Db	906	GTCTACACTCTGGCTCCAGCTATGCTCTGATCCAAAGCAAGGTGACAGAACTCCAG	965
Qy	901	CTTGGGCTGCTTCCACCCAG	924
Db	966	CTCTGTGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG	989

RESULT 4
US-10-051-874-25
; Sequence 25, Application US/10051874
; Publication No. US20040005557A1
; GENERAL INFORMATION:
; APPLICANT: Padigaru, Muralidhara
; APPLICANT: Alsobrook II, John P
; APPLICANT: Colman, Steven D
; APPLICANT: Spytek, Kimberly A
; APPLICANT: Boldos, Ferenc AM
; APPLICANT: Vernet, Corine AM
; APPLICANT: Li, Li
; APPLICANT: Shenoy, Suresh G
; APPLICANT: Casman, Stacie J
; APPLICANT: Guo, Xiaojia Saba
; APPLICANT: Edinger, Shlomit R
; APPLICANT: Macdougall, John R
; APPLICANT: Malyankar, Uriel M
; APPLICANT: Patturajan, Meera
; APPLICANT: Shimkets, Richard A
; APPLICANT: Pena, Carol BA
; APPLICANT: Tchernev, Velizar T
; APPLICANT: Zertusen, Bryan D
; APPLICANT: Millet, Isabelle
; APPLICANT: Miller, Charles E
; APPLICANT: Lepley, Denise M
; APPLICANT: Smithson, Glennda
; APPLICANT: Baumgartner, Jason C
; APPLICANT: Heriman, John L
; APPLICANT: Peyman, John A
; APPLICANT: Gorman, Linda
; APPLICANT: Mezes, Peter D
; APPLICANT: Kekuda, Ramesh
; APPLICANT: Taupier Jr, Raymond J
; APPLICANT: Gerlach, Valerie
; APPLICANT: Grosse, William M
; APPLICANT: Liu, Xiaohong
; APPLICANT: Ellerman, Karen
; APPLICANT: Rothenberg, Mark
; APPLICANT: Stone, David J
; APPLICANT: Burgess, Catherine E
; TITLE OF INVENTION: PROTEINS, POLYNUCLEOTIDES ENCODING THEM AND METHODS OF
; FILE REFERENCE: 21402-245
; CURRENT APPLICATION NUMBER: US/10/051,874
; CURRENT FILING DATE: 2002-09-25
; PRIOR APPLICATION NUMBER: 60/268,595
; PRIOR FILING DATE: 2001-02-14
; PRIOR APPLICATION NUMBER: 60/325,306
; PRIOR FILING DATE: 2001-09-27
; PRIOR APPLICATION NUMBER: 60/262,587
; PRIOR FILING DATE: 2001-01-18
; PRIOR APPLICATION NUMBER: 60/272,409
; PRIOR FILING DATE: 2001-02-28
; PRIOR APPLICATION NUMBER: 60/262,454
; PRIOR FILING DATE: 2001-01-18
; PRIOR APPLICATION NUMBER: 60/276,777
; PRIOR FILING DATE: 2001-03-15
; PRIOR APPLICATION NUMBER: 60/291,672
; PRIOR FILING DATE: 2001-05-17
; PRIOR APPLICATION NUMBER: 60/330,336
; PRIOR FILING DATE: 2001-10-18
; PRIOR APPLICATION NUMBER: 60/265,530
; PRIOR FILING DATE: 2001-01-31
; PRIOR APPLICATION NUMBER: 60/261,376

; PRIOR FILING DATE: 2001-01-16
 ; NUMBER OF SEQ ID NOS: 269
 ; SOFTWARE: PatentIn Ver. 2.1
 ; SEQ ID NO 25
 ; LENGTH: 1020
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 US-10-051-874-25

Query Match 20.5%; Score 225.6; DB 15; Length 1020;
 Best Local Similarity 58.0%; Pred. No. 1.3e-49;
 Matches 443; Conservative 0; Mismatches 309; Indels 12; Gaps 2;

QY	129	CTGCGGGCGCCCTGAGCCCTCGGCCCGCATCGTGGGGGCTCAAACGCGCAGCGCGGCAC	188
DB	171	CTGCGGGCAGCCCGCATGTCAGTCGGATCGTGGGGCCGGGATGGCGGACGAGA	230
QY	189	CTGGCCCTTGGCAAGTGAGCTGACCATGAGGTGGCCACATCTGGGGGGCTCCCTCAT	248
DB	231	GTGGCCGTGGCAGCGAGCATCAGCATGTGGGGCACAGTGTGGGGGGTGGCTCAT	290
QY	249	CGCCCTCTCTGGGCTCTTCCCTGCTCACTGTTTCATGAAGATGGACGTTGGAGCC	308
DB	291	CGCCCTCTCTGGGCTCTTCCCTGCTCACTGTTTCATGAAGATGGACGTTGGAGCC	341
QY	309	CGCGCCGAGTGTGGTACTGCTGGGCTGCTGCTCCAGGAGCGGCCCTTGGACGGCG	368
DB	342	GCAGGTGAGTACCGGTGCGCTGGGGGGCTGCTGGGCTCCACCTGCGCCCGCAC	401
QY	369	GCACACCGCGAGTGGCCCTCTGTTGGTGGCGGCACTACAGCCAAAGTGGAGCTGGG	428
DB	402	GCTCTGGTCCCGTGGCAGGGTGTGCTGCTCCCGGACTACTCCGAGGACGGGCCCG	461
QY	429	CGCCGACCTGGCCCTGTGGCTTGGCTTCCCGGAGCTGGGCTGGGCTGGGCTGGGCT	488
DB	462	CGCGCAGCTGGCCTGTGGCTTGGCTTGGCTTGGCTTGGCTTGGCTTGGCTTGGCT	521
QY	489	TGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT	548
DB	522	CGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT	591
QY	549	CTGGGAGAGCTCCAGAGGAGATCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT	608
DB	582	CTGGGAGAGCTCCAGAGGAGATCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT	641
QY	609	GCTAAGGCTGCTGGGAGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT	668
DB	642	GCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT	701
QY	669	CTTCACTCTCCAGATAT---TGCCAGGAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCT	725
DB	702	CCAGGCTGAGGCTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT	761
QY	726	GGACACCTGCGCAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT	785
DB	762	GGACGCTGCGCAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT	821
QY	786	CCAGGAGGATACACAGCTTGGGTTGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT	845
DB	822	CTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT	881
QY	846	CACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT	889
DB	882	CACAGTGTGGCCACATATAGCCCTGGATTCAGGCTCGCTCA	925

RESULT 5
 US-10-041-400A-1
 ; Sequence 1, Application US/10041400A
 ; Publication No. US20020110895A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Darrow, Andrew
 ; APPLICANT: Andrade-Gordon, Patricia

; APPLICANT: Qi, Jensen
 ; TITLE OF INVENTION: DNA Encoding the Human Serine
 ; FILE OF INVENTION: Protease EOS
 ; FILE REFERENCE: ORT-1031
 ; CURRENT APPLICATION NUMBER: US/10/041,400A
 ; CURRENT FILING DATE: 2002-01-08
 ; PRIOR APPLICATION NUMBER: US/09/387,375
 ; PRIOR FILING DATE: 1999-08-31
 ; NUMBER OF SEQ ID NOS: 9
 ; SOFTWARE: PatentIn Ver. 2.0
 ; SEQ ID NO 1
 ; LENGTH: 1613
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 US-10-041-400A-1

Query Match 20.3%; Score 224; DB 13; Length 1613;
 Best Local Similarity 57.9%; Pred. No. 3.6e-49;
 Matches 442; Conservative 0; Mismatches 310; Indels 12; Gaps 2;

QY	129	CTGCGGGCGCCCTGAGCCCTCGGCCCGCATCGTGGGGGCTCAAACGCGCAGCGCGGCAC	188
DB	149	CTGCGGGCAGCCCGCATGTCAGTCGGATCGTGGGGCCGGGATGGCGGACGAGA	208
QY	189	CTGGCCCTTGGCAAGTGAGCTGACCATGAGGTGGCCACATCTGGGGGGCTCCCTCAT	248
DB	209	GTGGCCGTGGCAGGCGAGCATCCAGCATCTGGGGCACACGTTGTGGGGGGTGGCTCAT	268
QY	249	CGCCCTCTCTGGGCTCTTCCCTGCTCACTGTTTCATGACGAATGGACGTTGGAGCC	308
DB	269	CGCCCTCTCTGGGCTCTTCCCTGCTCACTGTTTCATGACGAATGGACGTTGGAGCC	319
QY	309	CGCGCCGAGTGTGGTACTGCTGGGCTGCTGCTCCAGGAGCGGCCCTTGGACGGCGC	368
DB	320	GCAGTGTGATACCGGTGCGCTGGGGGGCTGGCTGCTGGGCTCCACCTGCGCCCGCAC	379
QY	369	GCACACCGCGCAGTGGCCCTCTGTTGGTGGCGGCAACTACAGCCAAAGTGGAGCTGGG	428
DB	380	GCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT	439
QY	429	CGCGGACCTGGCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT	488
DB	440	CGCGGACCTGGCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT	499
QY	489	TGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT	548
DB	500	CGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT	559
QY	549	CTGGGAGAGCTCCAGAGGAGATCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT	608
DB	560	CTGGGAGAGCTCCAGAGGAGATCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT	619
QY	609	GCTAAGGCTGCTGGGAGGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT	668
DB	620	GGTGGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT	679
QY	669	CTTCACTCTCCAGATAT---TGCCAGGAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCT	725
DB	680	CCAGGCTGAGGCTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT	739
QY	726	GGACACCTGCGCAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT	785
DB	740	GGACGCTGCGCAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT	799
QY	786	CCAGGAGGATACACAGCTTGGGTTGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT	845
DB	800	CTTGTGGGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT	859
QY	846	CACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT	889
DB	860	CACAGTGTGGCCACATATAGCCCTGGATTCAGGCTCGCTCA	903

RESULT 6

US-10-041-264A-1
; Sequence 1, Application US/10041264A
; Publication No. US20020142446A1
; GENERAL INFORMATION:
; APPLICANT: Darrow, Andrew
; APPLICANT: Andrade-Gordon, Patricia
; APPLICANT: Qi, Jensen
; TITLE OF INVENTION: DNA Encoding the Human Serine
; TITLE OF INVENTION: Protease EOS
; FILE REFERENCE: ORT-1031
; CURRENT APPLICATION NUMBER: US/10/041,264A
; PRIOR FILING DATE: 2002-01-08
; PRIOR APPLICATION NUMBER: US/09/387,375
; PRIOR FILING DATE: 1999-08-31
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 1
; LENGTH: 1613
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-041-264A-1

Query Match 20.3%; Score 224; DB 13; Length 1613;
Best Local Similarity 57.9%; Pred. No. 3.6e-49;
Matches 442; Conservative 0; Mismatches 310; Indels 12; Gaps 2;

QY	129	CTGCGGGCGCCCTGAGCCCTCGGCCCGCATCGTGGGGGCTCAAACGCGCAGCCCGGCAC	188
DB	149	CTGCGGGCGAGCCCGCATGTCAGTCGGATCGTTGGGGCCCGGATGGCCGGACGGAGA	208
QY	189	CTGGCCCTTGGCAAGTAGAGCTGCACCATGAGGTGGCCACATCTGGGGGGCTCCCTCAT	248
DB	209	GTGGCCGTGGCAGGCGAGCATCAGCATCTGGGGCACAGTGTGGGGGGTTCGCTCAT	268
QY	249	CGCCCCCTCTCGGGTCTCTCCGCTGCTCACTGTTTCATGAAGATGGACGTTGGAGCC	308
DB	269	CGCCCCCAGTGGTGTCTGACAGCGCGCATCTGCTTCCCCAGGA-----GGGCACT	319
QY	309	CGCGCCGCGAGTGGTGGTACTGCTGGGGCTGCACATCCAGAGCAGGCGCCCTGGACGGCG	368
DB	320	GCAGCTGAGTACCGGGTGGCTGGGGGCGTGGTCTGGGGTCCACTCGCCCGCGAC	379
QY	369	GCACACCCCGCAGTGGCCGCACTCGTGTGCGGGCCAACTACAGCAAGTGGAGTGGG	428
DB	380	GCTCTGGTGGCCGTGGACGGGTGCTGTGCTGCCCGGACTACTCCGAGGACGGGGCCCG	439
QY	429	CGCGACCTGGGCCCTGTCGGCTTCAACCGCCAGCTGGGGCCCGCGCGCTGTGGCC	488
DB	440	CGCGGACCTGGCACTGCTGAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG	499
QY	489	TGCTGCTGCGCCCGCGCTCAACCGCTTGTGACGCGCACCGCTGTGGGGCCACCGG	548
DB	500	CGTGTGCTGCGCCCG	559
QY	549	CTGGGGAGACGTCCAGAGGACAGATCTCTGCTGCTCTCCCTTGGGTGTACAGGAAGTGA	608
DB	560	CTGGGGCAGCTCCGCGCCAGAGTGGCTCCCGAGAGTGGCGACCGCTACAAGAGTAAG	619
QY	609	GCTAAGGCTGTGGGGGAGGCCACCTGTCAATGTCTCTACAGCCAGCCCGGTCCCTTCAA	668
DB	620	GGTGGCCGCTGTGGACCTCGCGCACCTTCGACGGGCTCTACCAAGTGGGGCGGACGTGC	679
QY	669	CCTCACTCTCCAGATAT---TGCAGGGATGCTGTGTGCTGCTACCCAGAGGGCCGCGAG	725
DB	680	CCAGGCTGAGCGCATGTGCTGCTGGAGTGTGTGCTGCTGCTGCTGCTGCTGCTGCTG	739
QY	726	GGACACCTGCGAGGCTGCTTGGGGGGCCCTGCTGTGTGAGAGAGCGCGCGCTGGTT	785
DB	740	GGACGCTGCGAGGCTGATTTCTGGGGGACCTCTGACCTGCTGCTGCTGCTGCTGCTG	799
QY	786	CCAGGCGAGGAATCAACAGCTTTGGGTTGGTGTGTGTGTGTGTGTGTGTGTGTGTGT	845

DB	800	CCTGTGGGGCGTGGTGGAGCTGGGGCAAGGGTGTGTGCTGCCCAACCCGTCCAGGGGTCTA	859
QY	846	CACTGCTGTGGCTACCTATGAGGATGATGATGATGATGATGATGATGATGATGATGATG	889
DB	860	CACCAAGTGTGGCCACATATAGCCCTTGGATTTCAGGCTGCGGTCA	903

RESULT 7

US-10-042-091A-1
; Sequence 1, Application US/10042091A
; Publication No. US20020142447A1
; GENERAL INFORMATION:
; APPLICANT: Darrow, Andrew
; APPLICANT: Andrade-Gordon, Patricia
; APPLICANT: Qi, Jensen
; TITLE OF INVENTION: DNA Encoding the Human Serine
; TITLE OF INVENTION: Protease EOS
; FILE REFERENCE: ORT-1031
; CURRENT APPLICATION NUMBER: US/10/042,091A
; PRIOR FILING DATE: 2002-01-08
; PRIOR APPLICATION NUMBER: US/09/387,375
; PRIOR FILING DATE: 1999-08-31
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 1
; LENGTH: 1613
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-042-091A-1

Query Match 20.3%; Score 224; DB 13; Length 1613;
Best Local Similarity 57.9%; Pred. No. 3.6e-49;
Matches 442; Conservative 0; Mismatches 310; Indels 12; Gaps 2;

QY	129	CTGCGGGCGCCCTGAGCCCTCGGCCCGCATCGTGGGGGCTCAAACGCGCAGCCCGGCAC	188
DB	149	CTGCGGGCGAGCCCGCATGTCAGTCGGATCGTTGGGGCCCGGATGGCCGGACGGAGA	208
QY	189	CTGGCCCTTGGCAAGTAGAGCTGCACCATGAGGTGGCCACATCTGGGGGGCTCCCTCAT	248
DB	209	GTGGCCGTGGCAGGCGAGCATCAGCATCTGGGGCACAGTGTGGGGGGTTCGCTCAT	268
QY	249	CGCCCCCTCTCGGGTCTCTCCGCTGCTCACTGTTTCATGAAGATGGACGTTGGAGCC	308
DB	269	CGCCCCCAGTGGTGTCTGACAGCGCGCATCTGCTTCCCCAGGA-----GGGCACT	319
QY	309	CGCGCCGCGAGTGGTGGTACTGCTGGGGCTGCACATCCAGAGCAGGCGCCCTGGACGGCG	368
DB	320	GCAGCTGAGTACCGGGTGGCTGGGGGCGTGGTCTGGGGTCCACTCGCCCGCGAC	379
QY	369	GCACACCCCGCAGTGGCCGCACTCGTGTGCGGGCCAACTACAGCAAGTGGAGTGGG	428
DB	380	GCTCTGGTGGCCGTGGACGGGTGCTGTGCTGCCCGGACTACTCCGAGGACGGGGCCCG	439
QY	429	CGCGACCTGGGCCCTGTCGGCTTCAACCGCCAGCTGGGGCCCGCGCGCTGTGGCC	488
DB	440	CGCGGACCTGGCACTGCTGAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG	499
QY	489	TGCTGCTGCGCCCGCGCTCAACCGCTTGTGACGCGCACCGCTGTGGGGCCACCGG	548
DB	500	CGTGTGCTGCGCCCG	559
QY	549	CTGGGGAGACGTCCAGAGGACAGATCTCTGCTGCTCTCCCTTGGGTGTACAGGAAGTGA	608
DB	560	CTGGGGCAGCTCCGCGCCAGAGTGGCTCCCGAGAGTGGCGACCGCTACAAGAGTAAG	619
QY	609	GCTAAGGCTGTGGGGGAGGCCACCTGTCAATGTCTCTACAGCCAGCCCGGTCCCTTCAA	668
DB	620	GGTGGCCGCTGTGGACCTTCGACGGGCTCTACCAAGTGGGGCGGACGTGC	679
QY	669	CCTCACTCTCCAGATAT---TGCAGGGATGCTGTGTGCTGCTACCCAGAGGGCCGCGAG	725
DB	680	CCAGGCTGAGCGCATGTGCTGCTGGAGTGTGTGCTGCTGCTGCTGCTGCTGCTGCTG	739

QY 726 GGACCTGCGAGGAGTCTCTGGGGGCCCCCTGCTCTGTGAGGAAGGCGCGCTGGTT 785
Db 740 GGACGCTGCGAGGGGATTTCTGGGGAACCTCTGACCTGCTGAGTCTGGAGCTGGGT 799
QY 786 CCAGGAGGAATCACCAGCTTTTGGGTTTGGCTGTGAAGGAGAAACCGCCCTGGAGTTT 845
Db 800 CTTGGTGGGCTGTGAGCTGGGGCAGGGTTGTGCTTCCCAACCGTCCAGGGTCTA 859
QY 846 CATGTGTGGCTACCTATCAGGATGATACCGGAGCAGGTGA 889
Db 860 CACCAGTGTGGCCACATATAGCCCCCTGATTACAGGCTCGGTCA 903

RESULT 8
US-09-948-094-1
; Sequence 1, Application US/09948094
; Patent No. US200090625A1
; GENERAL INFORMATION:
; APPLICANT: The Brigham and Women's Hospital, Inc.
; APPLICANT: Mok, Samuel
; APPLICANT: Wong, Kwong-kwok
; TITLE OF INVENTION: Methods of Detecting Cancer Based on Prostatin
; FILE REFERENCE: 81994/282423
; CURRENT APPLICATION NUMBER: US/09/948,094
; CURRENT FILING DATE: 2001-09-07
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 1
; LENGTH: 1834
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (229)..(1260)
US-09-948-094-1

Query Match 20.1%; Score 221.2; DB 9; Length 1834;
Best Local Similarity 54.2%; Pred. No. 2e-48;
Matches 501; Conservative 0; Mismatches 408; Indels 15; Gaps 2;

QY 1 GGGCCCTTGTCTCTGGGCGCATGCCAGAGGGGTCTCTGGGGCTTGGGCGCTGGGGCT 60
Db 211 GGGGCGCTTGTCTCTGGGCGCATGCCAGAGGGGTCTCTGGGGCTTGGGCGCTGGGGCT 270
QY 61 GTGGCCAAATCTGACTCATCTCACTTTACGGGTGTGGTGGCGTCCGGACCGCTAGGGC 120
Db 271 GTGGCCAAATCTGACTCATCTCACTTTACGGGTGTGGTGGCGTCCGGACCGCTAGGGC 330
QY 121 CCCCCGTAATCTGGGCGCGCTTGGGCGCTTGGGCGCTTGGGCGCTTGGGCGCTTGGGCG 180
Db 331 GCTCCCTCGGTGTG-----GCCGCCAAGCAGCATCACAGTGGCGAGTGCAGTC 384
QY 181 CCGGCACTGGCCCTTGGCAATGAGCTGCACCATGAGTGGCCACATCTGGGGGCG 240
Db 385 GCGGTGATGGCCCTTGGCAATGAGCTGCACCATGAGTGGCCACATCTGGGGGCG 444
QY 241 TCCCTCATCGCCCGCTTGGGCGCTTGGGCGCTTGGGCGCTTGGGCGCTTGGGCGCTTGGGCG 300
Db 445 TCTCTGTGTGTGAGAGTGGGTGTGTGAGTGTGTGAGTGTGTGAGTGTGTGAGTGTGTGAGT 497
QY 301 TTGGAGCCCGCGCGAGTGGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGT 360
Db 498 --GCACCAACAAGAGAGCTTATGAGGTCAAGCTGGGGGCGCCACACAGTACTCTTACTCC 555
QY 361 GAGGGCGGCACACCGCGAGTGGCGCGCTTGGGCGCTTGGGCGCTTGGGCGCTTGGGCGCTTGGGCG 420
Db 556 GAGGAGCGCAAGTCAAGCTTGGGCGCTTGGGCGCTTGGGCGCTTGGGCGCTTGGGCGCTTGGGCG 615
QY 421 GAGTGGGGCGCGAGCTGGCGCTTGGGCGCTTGGGCGCTTGGGCGCTTGGGCGCTTGGGCGCTTGGGCG 480
Db 616 GGTCCAGGGGCAATTTGCACTTCTCCAACTCAGCAGACCATCACCTTCTCCCGCTAC 675

QY 481 GTGTGGCCCTGCTCTGCTGCCCGCCCTCACACCGCTTGTGACGGACCGGCTGCTGG 540
Db 676 ATCCGGCCCATCTGCTTCCCTGAGCAACGCTCTCTCCCAAGCGCTTCCACTGCACT 735
QY 541 GCACCGGCTGGGAGAGCTCCAGGAGGAGATCTCTGCTCTTCCCTGCTGCTGGGTGTGACAG 600
Db 736 GTCACTGGCTGGGCTCATGTGGCCCTCAGTGGAGCTCTGACGCGCCCAAGCCACTGCGAG 795
QY 601 GAAGTGGAGCTAAGGCTGTCTGGGCGAGGCCACCTGTCAATGTCTTACAGCCAGCCCGGT 660
Db 796 CAACTCGAGGTGCTCTGATCAGTGTGAGACGTGTAACTGCTGTACAACTCGACGCGC 855
QY 661 CCCTTCAACCTCACTCTCCAGATATTCAGGAGTCTGTGCTGGCTACCCAGAGGGC 720
Db 856 AAGCTTGAGGAGCGCCACTTTGTCCAAAGAGACATGTTGTGTCTGTGCTGTGAGGGG 915
QY 721 CGCAGGAGACACTGCCAGAGTGACTCTGGGGGCGCCCTGCTCTGTGAGGAAGCGCGCGC 780
Db 916 GGCAGAGAGCGCTGCCAGGGTGACTCTGGGGGCGCCACTCTCTGCTGCTGTGGAGGCTCTC 975
QY 781 TGGTTCCAGGCGAGGAATCACCAGCTTTGGCTTGGCTGTGGAGCGAGAAACCGCCCTTGA 840
Db 976 TGGTACTTACGCGGCAATTTGTAGCTGGGAGATGCTGTGGGGCGCCCAACAGGCGCTGT 1035
QY 841 GTTTTCACTGCTGTGCTACCTATGAGGATGATAGGAGGAGCAGGTGATGGTTTCAGAG 900
Db 1036 GTGTACACTCTGGCCCTCCAGCTATGCTCTCTGATCCAAAGCAAGGTGACAGAACTCCAG 1095
QY 901 CTGGGCGCTGCTCTTCCACCCAG 924
Db 1096 CTTGTGTGTGCGCCCAAGCCAG 1119

RESULT 9
US-09-880-107-2214
; Sequence 2214, Application US/09880107
; Patent No. US20020142981A1
; GENERAL INFORMATION:
; APPLICANT: Horne, Darci T.
; APPLICANT: Vockley, Joseph G.
; APPLICANT: Scherf, Uwe
; APPLICANT: Gene Logic, Inc.
; TITLE OF INVENTION: Gene Expression Profiles in Liver Cancer
; FILE REFERENCE: 44921-5028-WO
; CURRENT APPLICATION NUMBER: US/09/880,107
; CURRENT FILING DATE: 2001-06-14
; PRIOR APPLICATION NUMBER: US 60/211,379
; PRIOR FILING DATE: 2000-06-14
; PRIOR APPLICATION NUMBER: US 60/237,054
; PRIOR FILING DATE: 2000-10-02
; NUMBER OF SEQ ID NOS: 3950
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2214
; LENGTH: 1834
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: Genbank Accession No. US20020142981A1 L41351
US-09-880-107-2214

Query Match 20.1%; Score 221.2; DB 9; Length 1834;
Best Local Similarity 54.2%; Pred. No. 2e-48;
Matches 501; Conservative 0; Mismatches 408; Indels 15; Gaps 2;

QY 1 GGGCCCTTGTCTCTGGGCGCATGCCAGAGGGGTCTCTGGGGCTTGGGCGCTGGGGCT 60
Db 211 GGGGCGCTTGTCTCTGGGCGCATGCCAGAGGGGTCTCTGGGGCTTGGGCGCTGGGGCT 270
QY 61 GTGGCCAAATCTGACTCATCTCACTTTACGGGTGTGGTGGCGTCCGGACCGCTAGGGC 120
Db 271 GTGGCCAAATCTGACTCATCTCACTTTACGGGTGTGGTGGCGTCCGGACCGCTAGGGC 330
QY 121 CCCCCGTAATCTGGGCGCGCTTGGGCGCTTGGGCGCTTGGGCGCTTGGGCGCTTGGGCG 180
Db 331 GCTCCCTCGGTGTG-----GCCGCCAAGCAGCATCACAGTGGCGAGTGCAGTC 384
QY 181 CCGGCACTGGCCCTTGGCAATGAGCTGCACCATGAGTGGCCACATCTGGGGGCG 240
Db 385 GCGGTGATGGCCCTTGGCAATGAGCTGCACCATGAGTGGCCACATCTGGGGGCG 444
QY 241 TCCCTCATCGCCCGCTTGGGCGCTTGGGCGCTTGGGCGCTTGGGCGCTTGGGCGCTTGGGCG 300
Db 445 TCTCTGTGTGTGAGAGTGGGTGTGTGAGTGTGTGAGTGTGTGAGTGTGTGAGTGTGTGAGT 497
QY 301 TTGGAGCCCGCGCGAGTGGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGT 360
Db 498 --GCACCAACAAGAGAGCTTATGAGGTCAAGCTGGGGGCGCCACACAGTACTCTTACTCC 555
QY 361 GAGGGCGGCACACCGCGAGTGGCGCGCTTGGGCGCTTGGGCGCTTGGGCGCTTGGGCGCTTGGGCG 420
Db 556 GAGGAGCGCAAGTCAAGCTTGGGCGCTTGGGCGCTTGGGCGCTTGGGCGCTTGGGCGCTTGGGCG 615
QY 421 GAGTGGGGCGCGAGCTGGCGCTTGGGCGCTTGGGCGCTTGGGCGCTTGGGCGCTTGGGCGCTTGGGCG 480
Db 616 GGTCCAGGGGCAATTTGCACTTCTCCAACTCAGCAGACCATCACCTTCTCCCGCTAC 675


```

Db      1096 CCTCGTGTGTGTCGCCCCCAACCCAG 1119
|||||
RESULT 11
US-10-097-340-261
; Sequence 261, Application US/10097340
; Publication No. US20030087250A1
; GENERAL INFORMATION:
; APPLICANT: John MONAHAN
; APPLICANT: Manjula GANNAVAPU
; APPLICANT: Sebastian HOERSCH
; APPLICANT: Shubhangi KAWATKAR
; APPLICANT: Steve G. KOVATS
; APPLICANT: Rachel E. MEYERS
; APPLICANT: Michael MORRISEY
; APPLICANT: Peter OLANDT
; APPLICANT: Ami SEN
; APPLICANT: Karen GLATT
; APPLICANT: Peter VEIBY
; APPLICANT: Gordon B. MILLS
; APPLICANT: Robert C. BAST, Jr.
; APPLICANT: Karen LU
; APPLICANT: Rosemarie SCHMANDT
; APPLICANT: Xumei ZHAO
; TITLE OF INVENTION: Nucleic Acid Molecules and Proteins For The Identification,
; FILE REFERENCE: MRI-030
; CURRENT APPLICATION NUMBER: US/10/097,340
; CURRENT FILING DATE: 2002-03-14
; PRIOR APPLICATION NUMBER: 60/276,025
; PRIOR FILING DATE: 2001-03-14
; PRIOR APPLICATION NUMBER: 60/325,149
; PRIOR FILING DATE: 2001-09-26
; PRIOR APPLICATION NUMBER: 60/276,026
; PRIOR FILING DATE: 2001-03-14
; PRIOR APPLICATION NUMBER: 60/324,967
; PRIOR FILING DATE: 2001/09/26
; PRIOR APPLICATION NUMBER: 60/311,732
; PRIOR FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: 60/325,102
; PRIOR FILING DATE: 2001-09-26
; PRIOR APPLICATION NUMBER: 60/323,580
; PRIOR FILING DATE: 2001-09-19
; NUMBER OF SEQ ID NOS: 363
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 261
; LENGTH: 1834
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-097-340-261

Query Match      20.1%; Score 221.2; DB 14; Length 1834;
Best Local Similarity 54.2%; Pred. No. 2e-48;
Matches 501; Conservative 0; Mismatches 408; Indels 15; Gaps 2;

Qy      1 GGGCCCTTGCTCTGGGCATGCCAGAGGGGTCTCTGGGCTGGGCAGCTGGGGCT 60
Db      211 GGGCCCTTGCTCTGGGCATGCCAGAGGGGTCTCTGGGCTGGGCAGCTGGGGCT 270

Qy      61 GTGGCCAACTGACTCATCTACTTACCGGTTGGTCCGTCGGACCCGCTAGGGGC 120
Db      271 GTGGCCAACTGCTCTATCTTTGGATTACTCCGGTGGGGACAGGAGCGAAGGGGAGAA 330

Qy      121 CCCCGTACTGGGGCGCCCTGAGCCCTCGGCCCTGCTGGGGGTCAAACGCGCAG 180
Db      331 GCTCCTCGGTGTG-----GCCCCCAAGCAGCATCAGGTGGCAGCTGCAGTC 384

Qy      181 CCGGCACCTGGCCCTGGCAAGTGAAGCTGACCATGAGGTGGCCACATCTGGGGGGC 240
Db      385 GCCGGTCAGTGGCCCTGGCAGTCAAGCATCACCATTGAAGCGCTCCATGTGTGTGGC 444

Qy      241 TCCCTCATCGCCCCCTCTCTGGGTCTCTCCGGTGTCTACTCTTTCATGACGAATGGGACG 300

```

```

RESULT 12
US-09-925-301-208
; Sequence 208, Application US/09925301
; Patent No. US20020052308A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA106
; CURRENT APPLICATION NUMBER: US/09/925,301
; CURRENT FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05982
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 1694
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 208
; LENGTH: 1668
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1505)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc_feature

```

```

Db      445 TCTCTGTGTCTGAGCAGTGGGTGCTGTGCTGCTCACTGTCTCCCAAGCGA----- 497
Qy      301 TTGGAGCCCGCGCGGAGTGGTGGTCTGCTGGGGGTGCTACTCCCGAGAGCGGGCCCTTG 360
Db      498 --GCACCAAGAGAGCCTATGAGTCAAGCTGGGGGCCCAACAGCTAGCTCTTACTCC 555
Qy      361 GAGGGCGGCACACCCGCGCAGTGGCGCCCATGTGTGGTCCCGGCCAACTACAGCCAAAGTG 420
Db      556 GAGGAGCGCAAGGTGAGCACCCTGAAGGACATCATCCCCACCCCCAGCTACTCCAGGAG 615
Qy      421 GAGTGGGGCGCGACCTGCTGCTGGCTGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 480
Db      616 GGTCTCCAGGGCGACATTCGACTCTCCAACTCAGCAGACCCATACCTTTCTCCCGCTAC 675
Qy      481 GTGTGGCTGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 540
Db      676 ATCCGGCCCATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 735
Qy      541 GCCACCGGTGGGGAGACGTCAGGAGGAGATTCCTCTGCTGCTGCTGCTGCTGCTGCTGCTG 600
Db      736 GTCACTGGCTGGGTCTATGTGCCCCCTCAGTGAGCTCTGACGCGCCCAAGCCACTGCAG 795
Qy      601 GAAGTGGAGCTAAGTCTGCTGGCGGAGGCCACTGTCATGCTCTACAGCCAGCCCGGT 660
Db      796 CAACTCGAGGTGCTCTGATCAGTGTGAGAGCGTGAATGCTGTGCTGTGCTGTGCTGTG 855
Qy      661 CCCTTCAACCTCACTCTCCAGATATTGCCAGGAGATCTGTGCTGCTGCTGCTGCTGCTG 720
Db      856 AAGCCTGAGGAGCGGCACATTGTCGAAGAGGACATGCTGTGCTGTGCTGTGCTGTGCTG 915
Qy      721 CGCAGGAGACATGCGCAGGGTGAATCTCTGGGGGCCCCCTGCTGTGAGGAGAGCGGGCCG 780
Db      916 GGCAGGAGCGCTGCGCAGGGTGAATCTCTGGGGGCCCCCTGCTGTGAGGAGAGCGGGTCTC 975
Qy      781 TGGTTCAGCGCAGGAATCACCAGCTTTGGCTTTGGCTGTGCTGCTGCTGCTGCTGCTGCTG 840
Db      976 TGGTACCTGACCGGCATTTGAGCTGGGAGATGCTGTGGGGCCCCGCAACAGGCTTGGT 1035
Qy      841 GTTTTTCACCTGTGCTACCTATGAGGCATGATACGGGAGCAGGTGATGGGTTTCAGAG 900
Db      1036 GTGTACACTCTGCGCTCCAGCTATGCTCTCTGATCCAAAGCAAGGTGACAGAACTCCAG 1095
Qy      901 CTGGGCGCTGCTCTTCCACCCAG 924
Db      1096 CCTCGTGTGTGTCGCCCAACCCAG 1119

```

LOCATION: (1565)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc.feature
LOCATION: (1598)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc.feature
LOCATION: (1620)
OTHER INFORMATION: n equals a,t,g, or c
US-09-925-301-208

Query Match 20.0%; Score 220.6; DB 9; Length 1668;
Best Local Similarity 54.7%; Pred. No. 2.8e-48;
Matches 506; Conservative 2; Mismatches 401; Indels 16; Gaps 3;
QY 1 GGGCCCTTCTCTGGCCATGCGCCAGAGGGGTCTCTGGGCTGGGCGCTGGGGCT 60
DB 191 GGGCCCTTCTCTGGCCATGCGCCAGAGGGGTCTCTGGGCTGGGCGCTGGGGCT 250
QY 61 GTGGCCAAATCTGACTCATCTACTTTACGGGTGGTGGCTCGGACCCCGCTAGGGGC 120
DB 251 GTGGCCAAATCTGCTCTATCTTGGATTACTCCGCTGGGACAGGACGCGAAGGGGCAGAA 310
QY 121 CCCCCGTACTGGGGCGCCCTGAGCCCTCGGCCCGCATCTGGGGGGCTCAACAGCGCAG 180
DB 311 GTCYCTCGGGTGTG-----GCCGCCAAGCACGCATCACAGGTGGCAGCTGCAGTTC 364
QY 181 CCGGGCACCTGGCCCTTGGCAAGTGAAGCTGCGACATGAGGTGGCCACATCTCGGGGGC 240
DB 365 GCGGTGACTGGCCCTGGAGGTGAGCATCACTATGAAGGGTCACTGTGTGGTGGC 424
QY 241 TCCCTCATCGCCCGCTCTCGGGTCTCTCGGTGTCTCACTGTTCATGACGAATGGGACG 300
DB 425 TCTCTGCTGTCTGAGCAGTGGGTGTCTGAGTGTCTCACTGTCTCCCGACGA----- 477
QY 301 TTGGAGCCCGCGCCGAGTGGTCTGCTGCTGGGGGTGCACTCCAGACGGGCCCTG 360
DB 478 --GCACCAAGGAAGCCATGAGGTCAAGTCTGGGGGCCCAAGTACTCTACTCTCC 535
QY 361 GACGGCGCGCACACCCGCGAGTGGCGCGCATCTGGTGGCGCGCATCTCCGCCCACTACAGCCAAAGTG 420
DB 536 GAGGAGCCCAAGGTGAGCACCTTGAAGGACATCATCCCCCAACCCAGCTTACCTCAGGAG 595
QY 421 GAGTGGGCGCGCACTGCGCCCTGTGCGCCCTGCTCCAGCCAGCTGGGCCCGCCCGC 480
DB 596 GGCTCCAGCGCGCACATGCACTCTCTCCAACTCAGCAGACCCATCACTCTCTCCCGTAC 655
QY 481 GTGTGGCTGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 540
DB 656 ATCCGGCCCATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 715
QY 541 GCCACCGGTGGGGAGACGTCCAGGAGGAGACCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 600
DB 716 GTCACTGGCTGGGGTCACTGTGGGCCCTCAGTGAGCTCTGAGCGCCCAAGCCACTGGAG 775
QY 601 GAAAGTGAGCTAAGGTGCTGGGCGAGGCCAAGCTGTCAA--TGTCTTCAAGCAGCGCCCG 659
DB 776 CAACTCGAGGTGCTCTGATCACTGCTGAGACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 835
QY 660 TCCCTTCAACCTCACTCTCCAGATATTGCCAGGATGCTGTGTGCTGGTATCCAGAGGG 719
DB 836 CAAAGCTGAGGAGCCGACCTTTGTCAAAGAGGACATGGTGTGTGCTGGGTATGTGAGGG 895
QY 720 CCGCAGGAGCACCTGCGAGGGTGACTCTGGGGGGCCCTGGTGTGTGAGGAAGCGGGCG 779
DB 896 GGGCAGGAGCGCTGCGAGGGTGACTCTGGGGGCCCACTCTCTGCTGCTGCTGCTGAGGGTCT 955
QY 780 CTGGTTCAGGAGGAATCACCAGCTTTTGGGTGTGGTGTGGACGGAGAAACCGCCCTGG 839
DB 956 CTGGTACCTGACGGGCAATTGTGAGCTGGGAGATGCTGTGGGGCCCGCAACAGGGCTGG 1015
QY 840 AGTTTTCAGTGTGTGGCTACCTATGAGGATGATGAGGACAGGTGATGGGTTCAGA 899
DB 1016 TGTGTACACTCTGGCTTCCAGCTTATGCTCTGATCCAAAGCAAGGTGACAGAACTCCA 1075

QY 900 GCCTGGGCTGCTTCCACCCAG 924
DB 1076 GCCTGCTGGTGGCCCAACCCAG 1100

RESULT 13

US-10-041-400A-8
Sequence 8, Application US/10041400A
Publication No. US20020110895A1
GENERAL INFORMATION:
APPLICANT: Darrow, Andrew
APPLICANT: Andrade-Gordon, Patricia
APPLICANT: Qi, Jenson
TITLE OF INVENTION: DNA Encoding the Human Serine
TITLE OF INVENTION: Protease EOS
FILE REFERENCE: ORT-1031
CURRENT APPLICATION NUMBER: US/10/041,400A
CURRENT FILING DATE: 2002-01-08
PRIOR APPLICATION NUMBER: US/09/387,375
PRIOR FILING DATE: 1999-08-31
NUMBER OF SEQ ID NOS: 9
SOFTWARE: Patentin Ver. 2.0
SEQ ID NO 8
LENGTH: 1130
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: Nucleic acid
OTHER INFORMATION: sequence of EOS zymogen fusion gene
US-10-041-400A-8

Query Match 19.7%; Score 216.8; DB 13; Length 1130;
Best Local Similarity 58.0%; Pred. No. 2.7e-47;
Matches 427; Conservative 0; Mismatches 297; Indels 12; Gaps 2;

QY 157 ATCTGGGGGGCTCAACGCGCAGCGGGGCACCTGGCCCTTGGCAAGTGAAGCTGACCAT 216
DB 166 ATCTGGGGGGCTATGCTCTAGAGGACGAGAGTGGCGTGGCAGGCGAGCATCCAGCAT 225
QY 217 GAGTGGGCGCATCTGGGGGGCTCCCTCATCGGCCCTCTCTGGGTCTCTCCGTGCT 276
DB 226 CTTGGGGCACAGTGTGGGGGGGTCTGCTCATCGGCCCTCTGAGTGGGTGCTGACAGCGCG 285
QY 277 CACTGTTTCATACCAATGGGACGTTGGAGCGCGCGCGAGTGGTCTGCTGCTGCTGCTG 336
DB 286 CACTGTTTCCCGAGG-----GGGACCTGCAGCTAGTACCGGTGGCTGGGG 336
QY 337 GTGCACTCCAGGACGGGCCCTCTGACCGCGCGCACACCGCGCAGTGGCCGCCATCGTG 396
DB 337 GGGCTGGCTCTGGGCTCCACCTCGGCCCGCACGCTCTCGGTGCCGTGCGAGGGTGTG 396
QY 397 GTGGCGCCCACTACAGCAAGTGAAGTGGGCGCGGACCTGGCCCTGCTGGCGCTGGCC 456
DB 397 CTGGCCCCCGGACTATCCGAGGACGGGGCCCGCGGACCTGGGCACTGCTGCTGAGTGG 456
QY 457 TCACCGCGCAGCTGGGGCCCGCGCTGTGGCTGTCTGCTCTCCCGCGCTCAACACCGC 516
DB 457 CGCCCGGTGGCCCTGAGCGCTCGCTCCAAACCGCTCTGCTGCTGCTGCTGCTGCTGCTG 516
QY 517 TTCTGTGACGGCACCGCTGTGGGCCACCGGCTGGGAGAGCTCCAGGAGGACATCT 576
DB 517 CGGCGCCCGGCACACCATGCGGGTCAACCGGCTGGGGCAGGCTCCGCGCAGAGTGGCC 576
QY 577 CTGGCTCTCCCTGGGTGTACAGGAAGTGAAGTAAAGGTTGCGGCTGCTGCTGCTGCTGCTG 636
DB 577 CTCCAGAGTGGCGACCCCTACAGGAGTAAAGGTTGCGGCTGCTGCTGCTGCTGCTGCTG 636
QY 637 CAATGTCTCTACAGCAGCGCGGTCCCTTCAACCTCACTCTCCAGATAT---TGCCAGGG 693
DB 637 GACGCGCTCTACCACTGGGCGCGGAGCTGCCCCAGGCTGAGCGCATTTGTGCTGCTGGG 696
QY 694 ATGCTGTGTGCTGGCTACCCAGAGGGGCCAGGGGACACTGTGCCAGGGTGACTCTGGGGGG 753


```
Db 697 AGTCTGTGTGCGGGTACCCAGGGGCCAAGAGACGCTTGCAGGGTGATTTCTGGGGGA 756
QY 754 CCCTGGTGTGTAGGAAGCGCCGCTGTTCACGAGCAGGAATACACAGCTTTGGGTTT 813
Db 757 CCTCTGACCTGCCTGCAGTCTGGAGCTGGGTCTCTGGTGGCGTGTGAGCTGGGGCAAG 816
QY 814 GGCTGTGGACGAGAAACCCCTCGGAGTTTTCACCTGTCTGTGGCTACCTATGAGGCATGG 873
Db 817 GGTGTGCTTGTCCCAACCTCCAGGGTCTACACAGTGTGGCCACATATAGCCCTGG 876
QY 874 ATACGGAGCAGGTGA 889
Db 877 ATTACGGCTCGCGTCA 892

RESULT 14
US-10-041-264A-8
; Sequence 8, Application US/10041264A
; Publication No. US2002014246A1
; GENERAL INFORMATION:
; APPLICANT: Darrow, Andrew
; APPLICANT: Andrade-Gordon, Patricia
; APPLICANT: Qi, Jensen
; TITLE OF INVENTION: DNA Encoding the Human Serine
; FILE REFERENCE: ORT-1031
; CURRENT APPLICATION NUMBER: US/10/041,264A
; CURRENT FILING DATE: 2002-01-08
; PRIOR APPLICATION NUMBER: US/09/387,375
; PRIOR FILING DATE: 1999-08-31
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 8
; LENGTH: 1130
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Nucleic acid
; OTHER INFORMATION: sequence of EOS zymogen fusion gene
US-10-041-264A-8
Query Match 19.7%; Score 216.8; DB 13; Length 1130;
Best Local Similarity 58.0%; Pred. No. 2,7e-47;
Matches 427; Conservative 0; Mismatches 297; Indels 12; Gaps 2;
QY 157 ATCGTGGGGGCTCAACGCGCAGCCGGCACCTGGCTTGGCAGTGGAGCTGCACCAT 216
Db 166 ATCGTTGGGGGCTATGCTCTAGAGGACGGAGTGGCCGTGGCAGGCGATCCAGCAT 225
QY 217 GGAGTGGGCGTCACTGTGGGGGGCTCCCTCATCGCCCCCTCTCTGGGTCTCTCCGTGCT 276
Db 226 CCTTGGGGCACAACGCGCAGCGGCGTCTAGAGGACGGAGTGGCCGTGGCAGGAGCATCCAGCAT 225
QY 277 GAGGTGGGCGCACAATGTGGGGGGCTCCCTCATCGCCCCCTCTCTGGGTCTCTCCGTGCT 276
Db 286 CACTGTCTTCCCGA-----GGGCACTGCGACGTAGTACCGCTGCGCGGCGGCGG 336
QY 337 GTGCACTTCCAGAACGCGGCGCTTGGAGCGCGCGCGAGTGGTCTGCTTGTCTGGGC 336
Db 337 GCGTGTGCTTGGGCTCCACCTCGCCCCCGCAGCGCTCTCGGTGCGCGGCGGCGG 396
QY 457 TCAACCGCGAGCTGGGCGCGCGTGTGGCTGTCTGTGCGTGGCGCGGCTTCAACCGC 516
Db 457 CGCCCGGTGCGGCTGAGGCGTGGCGTCAACCGCGTCTGCGTGGCGCGGCGGCGG 516
QY 517 TTCGTGACGCGACCGCTGTGGGCGACCGGCTGGGGAGAGCTGCGAGGAGGAGATCCT 576
Db 517 CGCGCGCGCGGCAACCATGCGGGTCAACCGGCTGGGGCGAGCCTTCCCGCGAGGATGCC 576
```

```
QY 577 CTGCTCTCCCTGGTGTCTACAGGAAGTGGAGCTAAAGGCTGCTGGGCGAGGCGACCTGT 636
Db 577 CTCACAGAGTGGGACCGGCTACAGAGGTAAGGGTGGCGCTGCTGAGACTCGCGCACCTGC 636
QY 637 CAATGTCTCTACAGCAGCCCGTCCCTTCACACCTCACTCTCCAGATAF---TGCCAGGG 693
Db 637 GACGGGCTCTACACAGTGGGCGCGGACGTGCCAGGCTGAGCGCATTTGTGCTGCGCTGGG 696
QY 694 ATGCTGTGTGCTGGCTACCCAGAGGGCGCAGAGGACACCTGCGCAGGTCAGTCTGGGGGG 753
Db 697 AGTCTGTGTGCGGGCTACCCCGAGGGCCACAAGGACGCGCTGCCAGGGTGAATTTCTGGGGGA 756
QY 754 CCCCTGGTCTGTGAGGAAGGCGCGCGCTGGTTCCAGGCGAGGAATCACACAGCTTTTGGGTTT 813
Db 757 CCTCTGACCTGCCTGCAGTCTGGAGCTGGGTCTCTGGTGGCGTGTGAGCTGGGGCAAG 816
QY 814 GGCTGTGGACGAGAAACCGCCCTCGAGTTTTCACCTGTGTGGCTACCTATGAGGCATGG 873
Db 817 GGTGTGCGCTGCCAACCGTCCAGGGGTCTACACAGTGTGGCCACATATAGCCCTGG 876
QY 874 ATACGGGACGAGTGA 889
Db 877 ATTACGGCTCGCGTCA 892

RESULT 15
US-10-042-091A-8
; Sequence 8, Application US/10042091A
; Publication No. US20020142447A1
; GENERAL INFORMATION:
; APPLICANT: Darrow, Andrew
; APPLICANT: Andrade-Gordon, Patricia
; APPLICANT: Qi, Jensen
; TITLE OF INVENTION: DNA Encoding the Human Serine
; FILE REFERENCE: ORT-1031
; CURRENT APPLICATION NUMBER: US/10/042,091A
; CURRENT FILING DATE: 2002-01-08
; PRIOR APPLICATION NUMBER: US/09/387,375
; PRIOR FILING DATE: 1999-08-31
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 8
; LENGTH: 1130
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Nucleic acid
; OTHER INFORMATION: sequence of EOS zymogen fusion gene
US-10-042-091A-8
Query Match 19.7%; Score 216.8; DB 13; Length 1130;
Best Local Similarity 58.0%; Pred. No. 2,7e-47;
Matches 427; Conservative 0; Mismatches 297; Indels 12; Gaps 2;
QY 157 ATCGTGGGGGCTCAACGCGCAGCCGGCACCTGGCTTGGCAGTGGAGCTGCACCAT 216
Db 166 ATCGTTGGGGGCTATGCTCTAGAGGACGGAGTGGCCGTGGCAGGAGCATCCAGCAT 225
QY 217 GAGGTGGGCGCACAATGTGGGGGGCTCCCTCATCGCCCCCTCTCTGGGTCTCTCCGTGCT 276
Db 226 CCTTGGGGCACAACGCGCAGCGGCGTCTAGAGGACGGAGTGGCCGTGGCAGGAGCATCCAGCAT 225
QY 277 GAGGTGGGCGCACAATGTGGGGGGCTCCCTCATCGCCCCCTCTCTGGGTCTCTCCGTGCT 276
Db 286 CACTGTCTTCCCGA-----GGGCACTGCGACGTAGTACCGCTGCGCGGCGGCGG 336
QY 337 GTGCACTTCCAGAACGCGGCGCTTGGAGCGCGCGCGAGTGGTCTGCTTGTCTGGGC 336
Db 337 GCGTGTGCTTGGGCTCCACCTCGCCCCCGCAGCGCTCTCGGTGCGCGGCGGCGG 396
QY 397 GTGCCGCGCGGCAACCATGCGGGTCAACCGGCTGGGGCGAGCCTTCCCGCGAGGATGCC 456
```

Db	397	CTGCCCCCGGACTACTCCGAGGACGGGGCCCGCGCGACCTGGCACTGCTGCAGCTGGT	456
Qy	457	TCACCCGCGCAGCCTGGGCGCCGCGGTGTTGGCTGTCTGCTGCGCCCGCGCTCACACGC	516
Db	457	CGCCCGGTGCCCCCTGAGGGCTCGGGTCCAAACCCGTCTGCTGCGCGCTGCGCGCGCGC	516
Qy	517	TTCGTGCACGGCACCGGCTGTGGGGCCACCGGCTGGGGAGACGTCCAGGAGGCGAGATCCT	576
Db	517	CCGCGCGCCGCGCACACCATGCCGGTACCGGCTGGGGCAGCCTCCGCGCAGGAGTGCCC	576
Qy	577	CTGCGCTCTCCCGCTGAGGAGTGAAGCTAAGGCTGCTGGGGGAGGCCACCTGT	636
Db	577	CTCCAGAGTGGCGACCGCTACAAGGAGTAAGGGTGGCGCTGCTGGACTCGCGCACCTGC	636
Qy	637	CAATGTCTCTACAGCACCGCGGTCCCTTCAACCTCACTCTCCAGATAT---TGCCAGGG	693
Db	637	GACGGCCTCTACAGGTGGGCGCGGACGTGCCCGAGGCTGAGCGCATGTGCTGCTGGG	696
Qy	694	ATGCTGTGTGCTGGCTACCCAGAGGGCGCGAGGACACCTGCCAGGGTGACTCTGGGGGG	753
Db	697	AGTCTGTGTGCGGCTTACCCCGAGGGCCACAAGGACGCTGCCAGGGTGATTCTGGGGGA	756
Qy	754	CCCCTGGTCTGTGAGGAGGGCGCGCTGGTTCCAGGCAGGAATCACAGCTTTGGGTTT	813
Db	757	CCTCTGACCTGCTGAGTCTGGGAGCTGGGTCTGCTGGGGCGTGGTGAGCTGGGGCAAG	816
Qy	814	GGCTGTGAGCGAGAAACCGCGCTGGAGTTTCACTGCTGTGGCTACCTATGAGGCATGG	873
Db	817	GGTGTGCTCCCGCCCAACCGTCCAGGGTCTACACCACTGTGGCCACATATAGCCCCTGG	876
Qy	874	ATACGGGAGCAGGTGA	889
Db	877	ATCAGGCTCGGTCA	892

Search completed: February 25, 2004, 12:28:25
Job time : 457 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 25, 2004, 11:19:55 ; Search time 4327 seconds
(without alignments)
11038.600 Million cell updates/sec

Title: US-10-037-417-45

Perfect score: 1102

Sequence: 1 gggcccttgctgggcat.....gggggtctgatggggcctcc 1102

Scoring table: OLIGO_NUC

Gapop 60.0 , Gapext 60.0

Searched: 3470272 seqs, 21671516995 residues

Word size : 0

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

GenEmbl.*

1: gb.ba.*

2: gb.htg.*

3: gb.in.*

4: gb.om.*

5: gb.ov.*

6: gb.pat.*

7: gb.ph.*

8: gb.pl.*

9: gb.pr.*

10: gb.ro.*

11: gb.sts.*

12: gb.sy.*

13: gb.un.*

14: gb.vi.*

15: em.ba.*

16: em.fun.*

17: em.hum.*

18: em.in.*

19: em.mu.*

20: em.om.*

21: em.or.*

22: em.ov.*

23: em.pat.*

24: em.ph.*

25: em.pl.*

26: em.ro.*

27: em.sts.*

28: em.un.*

29: em.vi.*

30: em.htg.hum.*

31: em.htg.inv.*

32: em.htg.other.*

33: em.htg.mus.*

34: em.htg.pln.*

35: em.htg.fod.*

36: em.htg.man.*

37: em.htg.vrt.*

38: em.sy.*

39: em.htgo.hum.*

40: em.htgo.mus.*

41: em.htgo.other.*

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	1102	100.0	1102	6	AX512287	AX512287 Sequence
2	1102	100.0	1102	6	AX512289	AX512289 Sequence
3	716	65.0	2457	6	AX360096	AX360096 Sequence
4	716	65.0	2681	6	AX480935	AX480935 Sequence
5	631	57.3	2810	6	BD127529	BD127529 Primer fo
6	631	57.3	2810	9	AK075142	AK075142 Homo sapi
7	486	44.1	537	6	AX342934	AX342934 Sequence
8	384	34.8	456	6	AX342936	AX342936 Sequence
9	346	31.4	670	6	BD125219	BD125219 Primer fo
10	270	24.5	127769	9	BD126485	BD126485 Primer fo
11	270	24.5	127769	2	AC009088	AC009088 Homo sapi
12	270	24.5	195476	9	AC135044	AC135044 Homo sapi
13	87	7.9	168064	9	AC093520	AC093520 Homo sapi
14	67	6.1	1796	6	AX098193	AX098193 Sequence
15	67	6.1	1800	9	BC001462	BC001462 Homo sapi
16	67	6.1	1835	6	AX098215	AX098215 Sequence
17	67	6.1	7008	9	HSU33446	U33446 Human prost
18	67	6.1	195476	2	AC135044	AC135044 Homo sapi
19	53	4.8	1726	6	AX675579	AX675579 Sequence
20	53	4.8	1834	6	AX335777	AX335777 Sequence
21	53	4.8	1834	6	AX336076	AX336076 Sequence
22	53	4.8	1834	6	AX336340	AX336340 Sequence
23	53	4.8	1834	6	AX409567	AX409567 Sequence
24	53	4.8	1834	6	AX474697	AX474697 Sequence
25	53	4.8	1834	9	HUMPROS	I41351 Homo sapien
26	51	4.6	596	6	AX193364	AX193364 Sequence
27	50	4.5	6207	9	BC035589	BC035589 Homo sapi
28	50	4.5	7604	9	AB002294	AB002294 Human mRN
29	50	4.5	182230	9	AC135050	AC135050 Homo sapi
30	49	4.4	1161	6	AX675581	AX675581 Sequence
31	35	3.2	64328	2	AC124816	AC124816 Mus muscu
32	35	3.2	71177	2	AC101465	AC101465 Mus muscu
33	35	3.2	185788	10	AC124461	AC124461 Mus muscu
34	35	3.2	219763	10	AC093175	AC093175 Mus muscu
35	35	3.2	232119	2	AC106629	AC106629 Rattus no
36	35	3.2	244161	2	AC117170	AC117170 Rattus no
37	26	2.4	26	6	AX512409	AX512409 Sequence
38	25	2.3	25	6	AX556574	AX556574 Sequence
39	25	2.3	28628	2	AC020159	AC020159 Drosophil
40	25	2.3	184272	3	AC009211	AC009211 Drosophil
41	25	2.3	221830	3	AC009393	AC009393 Drosophil
42	25	2.3	247955	3	AE003698	AE003698 Drosophil
43	24	2.2	402	6	AX369313	AX369313 Sequence
44	24	2.2	610	6	AR263844	AR263844 Sequence
45	24	2.2	995	4	DOGMCITRPA	M24664 Dog mast ce

ALIGNMENTS

RESULT 1
AX512287
LOCUS AX512287 1102 bp DNA linear PAT 27-SEP-2002
DEFINITION Sequence 43 from Patent WO02053742.
ACCESSION AX512287
VERSION AX512287.1 GI:23392631
KEYWORDS Homo sapiens (human)
SOURCE
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
AUTHORS
1
Kekuda,R., Alsobrook,J.P., Tchernev,V.T., Liu,X., Spytek,X.A.,
Patturajan,M., Grosse,W.M., Lepley,D.M., Burgess,C.E., Vernet,C.A.,
Li,L., Gorman,L., Edinger,S., Scioire,P., Ellerman,K., Malyankar,U.,

Pred. No. is the number of results predicted by chance to have a

Rothenberg, M., Stone, D., Boldog, F., Shenoy, S. and Anderson, D.
Proteins and nucleic acids encoding same
Patent: WO 02053742-A 43 11-JUL-2002;
Curagen Corporation (US)

FEATURES

source
1..1102
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"

ORIGIN

Query Match 100.0%; Score 1102; DB 6; Length 1102;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1102; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GGGCCCTTGTCTGGGCAATGCCAGAGGGGCTCTGGGGCTGGGACGCTGGGGGCT 60
DB 1 GGGCCCTTGTCTGGGCAATGCCAGAGGGGCTCTGGGGCTGGGACGCTGGGGGCT 60

QY 61 GTGGCCAAATCTGACTCATATCACTTTACGGGTGGTGCCGCTCCGGACCCGCTAGGGC 120
DB 61 GTGGCCAAATCTGACTCATATCACTTTACGGGTGGTGCCGCTCCGGACCCGCTAGGGC 120

QY 121 CCCCCGCTACTGGGGCCCTGAGCCCTCGCCCGCATCTGGGGGGCTCAAAACGGCGAG 180
DB 121 CCCCCGCTACTGGGGCCCTGAGCCCTCGCCCGCATCTGGGGGGCTCAAAACGGCGAG 180

QY 181 CCGGGACCTTGGCTTGGCAAGTGAGCTGCACCATGGAGGTGGCCACATCTCGGGGGC 240
DB 181 CCGGGACCTTGGCTTGGCAAGTGAGCTGCACCATGGAGGTGGCCACATCTCGGGGGC 240

QY 241 TCCTCATCGCCCTCTCGGTCTCTCGCTGCTCACTGTTCATGACAAATGGAGC 300
DB 241 TCCTCATCGCCCTCTCGGTCTCTCGCTGCTCACTGTTCATGACAAATGGAGC 300

QY 301 TTGGAGCCCGCGCGAGTGGTGGTCTGCTGGGCTGCACTCCACAGACGGGCCCTG 360
DB 301 TTGGAGCCCGCGCGAGTGGTGGTCTGCTGGGCTGCACTCCACAGACGGGCCCTG 360

QY 361 GACGGCGCGACACCCGCGAGTGGCCCATCTGGTGCCGCGCAACTACACCGCAAGTG 420
DB 361 GACGGCGCGACACCCGCGAGTGGCCCATCTGGTGCCGCGCAACTACACCGCAAGTG 420

QY 421 GAGTGGGCGCGACCTGGCCCTGCTGCGCTGGCTTCAACCGCCAGCTTGGCCCGCC 480
DB 421 GAGTGGGCGCGACCTGGCCCTGCTGCGCTGGCTTCAACCGCCAGCTTGGCCCGCC 480

QY 481 GTGTGGCTCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 540
DB 481 GTGTGGCTCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 540

QY 541 GCCACCGGCTGGGAGACGTCCAGGAGGCAATCTCTGCTCTCTCTCTCTCTCTCTCT 600
DB 541 GCCACCGGCTGGGAGACGTCCAGGAGGCAATCTCTGCTCTCTCTCTCTCTCTCTCT 600

QY 601 GAAGTGGAGTAAAGTGTCTGGGCGAGGCCAATCTGTAATGTCTCTACAGCCAGCCGGT 660
DB 601 GAAGTGGAGTAAAGTGTCTGGGCGAGGCCAATCTGTAATGTCTCTACAGCCAGCCGGT 660

QY 661 CCTTCAACCTCACTCTCCAGATATTGCCAGGATGCTGTGTGTGTGTGTGTGTGTGT 720
DB 661 CCTTCAACCTCACTCTCCAGATATTGCCAGGATGCTGTGTGTGTGTGTGTGTGTGT 720

QY 721 CGCAGGACACCTGCCAGGGTGAATCTGGGGGGCCCTGCTGTGTGTGTGTGTGTGTGTGT 780
DB 721 CGCAGGACACCTGCCAGGGTGAATCTGGGGGGCCCTGCTGTGTGTGTGTGTGTGTGTGT 780

QY 781 TGGTTCAGGAGGAATCAACAGCTTTGGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 840
DB 781 TGGTTCAGGAGGAATCAACAGCTTTGGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 840

QY 841 GTTTTCACTCTGTGGCTACTATGAGCATGATACGGGAGCAGGTGATGGGTTTCAGAG 900

DB 841 GTTTTCACTCTGTGGCTACTATGAGCATGGATAGGAGCAGGTGATGGTTTCAGAG 900

QY 901 CTGGGCTCTGCTTCCACCCAGCCAGAGACCCAGTCTGATTTTCAATCAAAACG 960
DB 901 CTGGGCTCTGCTTCCACCCAGCCAGAGACCCAGTCTGATTTTCAATCAAAACG 960

QY 961 GCAATCTCTGATTTGCCAGAAATCTTTTGGGCCCTTCTCCCATATATCAGTAGGAGTC 1020
DB 961 GCAATCTCTGATTTGCCAGAAATCTTTTGGGCCCTTCTCCCATATATCAGTAGGAGTC 1020

QY 1021 TCACTGGACCAAAAGCCTTCTCTCCCTCGCTCTCTCCACACTCTCTCTGGGCCCTC 1080
DB 1021 TCACTGGACCAAAAGCCTTCTCTCCCTCGCTCTCTCCACACTCTCTCTGGGCCCTC 1080

QY 1081 TGGGGGTTCTGATGGGGCTCC 1102
DB 1081 TGGGGGTTCTGATGGGGCTCC 1102

RESULT 2
AX512289
LOCUS 1102 bp DNA linear PAT 27-SEP-2002
DEFINITION Sequence 45 from Patent WO02053742.
ACCESSION AX512289
VERSION AX512289.1 GI:23392632
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1
Kekuda, R., Alsbrook, J. P., Tchernev, V. T., Liu, X., Spytek, K. A.,
Patturajan, M., Grosse, W. M., Lepley, D. M., Burgess, C. E., Vernet, C. A.,
Li, L., Gorman, L., Edinger, S., Sciore, P., Ellerman, K., Malyankar, U.,
Rothenberg, M., Stone, D., Boldog, F., Shenoy, S. and Anderson, D.
Proteins and nucleic acids encoding same
Patent: WO 02053742-A 45 11-JUL-2002;
Curagen Corporation (US)

LOCATION/Qualifiers
1..1102
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"

TITLE
JOURNAL
FEATURES
source
1..1102
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"

ORIGIN
Query Match 100.0%; Score 1102; DB 6; Length 1102;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1102; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GGGCCCTTGTCTGGGCAATGCCAGAGGGGCTCTGGGGCTGGGACGCTGGGGGCT 60
DB 1 GGGCCCTTGTCTGGGCAATGCCAGAGGGGCTCTGGGGCTGGGACGCTGGGGGCT 60

QY 61 GTGGCCAAATCTGACTCATATCACTTTACGGGTGGTGCCGCTCCGGACCCGCTAGGGC 120
DB 61 GTGGCCAAATCTGACTCATATCACTTTACGGGTGGTGCCGCTCCGGACCCGCTAGGGC 120

QY 121 CCCCCGCTACTGGGGCCCTGAGCCCTCGCCCGCATCTGGGGGGCTCAAAACGGCGAG 180
DB 121 CCCCCGCTACTGGGGCCCTGAGCCCTCGCCCGCATCTGGGGGGCTCAAAACGGCGAG 180

QY 181 CCGGGACCTTGGCTTGGCAAGTGAGCTGCACCATGGAGGTGGCCACATCTCGGGGGC 240
DB 181 CCGGGACCTTGGCTTGGCAAGTGAGCTGCACCATGGAGGTGGCCACATCTCGGGGGC 240

QY 241 TCCTCATCGCCCTCTCGGTCTCTCGCTGCTCACTGTTCATGACAAATGGAGC 300
DB 241 TCCTCATCGCCCTCTCGGTCTCTCGCTGCTCACTGTTCATGACAAATGGAGC 300

QY 301 TTGGAGCCCGCGCGAGTGGTGGTCTGCTGGGCTGCACTCCACAGACGGGCCCTG 360
DB 301 TTGGAGCCCGCGCGAGTGGTGGTCTGCTGGGCTGCACTCCACAGACGGGCCCTG 360

QY 361 GACGGCGCGACACCCGCGAGTGGCCCATCTGGTGCCGCGCAACTACACCGCAAGTG 420
DB 361 GACGGCGCGACACCCGCGAGTGGCCCATCTGGTGCCGCGCAACTACACCGCAAGTG 420

QY 421 GAGTGGGCGCGACCTGGCCCTGCTGCGCTGGCTTCAACCGCCAGCTTGGCCCGCC 480
DB 421 GAGTGGGCGCGACCTGGCCCTGCTGCGCTGGCTTCAACCGCCAGCTTGGCCCGCC 480

QY 481 GTGTGGCTCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 540
DB 481 GTGTGGCTCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 540

QY 541 GCCACCGGCTGGGAGACGTCCAGGAGGCAATCTCTGCTCTCTCTCTCTCTCTCTCT 600
DB 541 GCCACCGGCTGGGAGACGTCCAGGAGGCAATCTCTGCTCTCTCTCTCTCTCTCTCT 600

QY 601 GAAGTGGAGTAAAGTGTCTGGGCGAGGCCAATCTGTAATGTCTCTACAGCCAGCCGGT 660
DB 601 GAAGTGGAGTAAAGTGTCTGGGCGAGGCCAATCTGTAATGTCTCTACAGCCAGCCGGT 660

QY 661 CCTTCAACCTCACTCTCCAGATATTGCCAGGATGCTGTGTGTGTGTGTGTGTGTGT 720
DB 661 CCTTCAACCTCACTCTCCAGATATTGCCAGGATGCTGTGTGTGTGTGTGTGTGTGT 720

QY 721 CGCAGGACACCTGCCAGGGTGAATCTGGGGGGCCCTGCTGTGTGTGTGTGTGTGTGTGT 780
DB 721 CGCAGGACACCTGCCAGGGTGAATCTGGGGGGCCCTGCTGTGTGTGTGTGTGTGTGTGT 780

QY 781 TGGTTCAGGAGGAATCAACAGCTTTGGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 840
DB 781 TGGTTCAGGAGGAATCAACAGCTTTGGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 840

QY 841 GTTTTCACTCTGTGGCTACTATGAGCATGATACGGGAGCAGGTGATGGGTTTCAGAG 900

Qy 361 GACGGCGCACACCCGCGCAGTGGCGCCATCGTGTGGCGGCCAACTACAGCCAAAGTG 420
Db 361 GACGGCGCACACCCGCGCAGTGGCGCCATCGTGTGGCGGCCAACTACAGCCAAAGTG 420
Qy 421 GAGTGGGGCGGACCTGGCCCTCTGCTGGCGCTGCGCTACCCGCGCAGCTGGGGCCCGCC 480
Db 421 GAGTGGGGCGGACCTGGCCCTCTGCTGGCGCTGCGCTACCCGCGCAGCTGGGGCCCGCC 480
Qy 481 GTGTGGCTGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 540
Db 481 GTGTGGCTGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 540
Qy 541 GCCACCGGCTGGGAGAGCTGCCAGGAGCGAGATCTCTGCTCTCCCTGGGTGCTACAG 600
Db 541 GCCACCGGCTGGGAGAGCTGCCAGGAGCGAGATCTCTGCTCTCCCTGGGTGCTACAG 600
Qy 601 GAAGTGAGCTAAGGCTGCTGGGCGAGGCGACCTGTCAATGTCTACAGCCAGCCCGGT 660
Db 601 GAAGTGAGCTAAGGCTGCTGGGCGAGGCGACCTGTCAATGTCTACAGCCAGCCCGGT 660
Qy 661 CCCTTCAACCTCACTCTCCAGATATGCCAGGAGTGTGTGCTGCTACCCAGAGGGC 720
Db 661 CCCTTCAACCTCACTCTCCAGATATGCCAGGAGTGTGTGCTGCTACCCAGAGGGC 720
Qy 721 CGCAGGAGACCTGCCAGGCTGCTCTGGGGGCGCCCTGCTGTGAGGAGGCGCGCGC 780
Db 721 CGCAGGAGACCTGCCAGGCTGCTCTGGGGGCGCCCTGCTGTGAGGAGGCGCGCGC 780
Qy 781 TGGTTCCAGGAGAGATCACAGCTTTGGGTTTGGCTGTGAGGAGGAGGAGGAGGAGGAG 840
Db 781 TGGTTCCAGGAGAGATCACAGCTTTGGGTTTGGCTGTGAGGAGGAGGAGGAGGAGGAG 840
Qy 841 GTTTCCTGCTGTGGCTTACCTATGAGCATGGATACGGGAGCAGGTGATGGGTTTCAGAG 900
Db 841 GTTTCCTGCTGTGGCTTACCTATGAGCATGGATACGGGAGCAGGTGATGGGTTTCAGAG 900
Qy 901 CCTGGGCGTGCCTTTCCACCCAGCCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 960
Db 901 CCTGGGCGTGCCTTTCCACCCAGCCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 960
Qy 961 GCATTCCTGAGATTCGCGAGATCCTTTTGGGCGCTTGTCCCATATATCAGTAGGATC 1020
Db 961 GCATTCCTGAGATTCGCGAGATCCTTTTGGGCGCTTGTCCCATATATCAGTAGGATC 1020
Qy 1021 TCAACTGGGACCAAAAGCTTGTCTCCCTCGGCTCTCTCCACACTCTCTCTGGGCGTC 1080
Db 1021 TCAACTGGGACCAAAAGCTTGTCTCCCTCGGCTCTCTCCACACTCTCTCTGGGCGTC 1080
Qy 1081 TGGGGGTTCTGATGGGCGCTCC 1102
Db 1081 TGGGGGTTCTGATGGGCGCTCC 1102

RESULT 3
AX360096
LOCUS AX360096 2457 bp DNA linear PAT 13-FEB-2002
DEFINITION Sequence 52 from Patent WO0200860.
ACCESSION AX360096
VERSION AX360096.1 GI:18675722
KEYWORDS Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1
AUTHORS Plowman, G., Whyte, D., Sudarsanam, S., Manning, G., Caenepeel, S. and Charyczak, G.
TITLE Novel proteases
JOURNAL Patent: WO 0200860-A 52 03-JAN-2002;
Sugen, Inc. (US)
FEATURES
1..2457 Location/Qualifiers
/organism="Homo sapiens"

ORIGIN
Query Match 65.0%; Score 716; DB 6; Length 2457;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 816; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
Qy 128 ACTGCGGGCGCCCTGAGCCCTCGGCGCGCATCTGCGGGGCTCAAAAGCGCAGCGCGGCA 187
Db 110 ACTGCGGGCGCCCTGAGCCCTCGGCGCGCATCTGCGGGGCTCAAAAGCGCAGCGCGGCA 169
Qy 188 CTTGGGCTTGGCAAGTGAGCTGCAACATGAGAGTGCCACATCTGCGGGGCTCCCTCA 247
Db 170 CTTGGGCTTGGCAAGTGAGCTGCAACATGAGAGTGCCACATCTGCGGGGCTCCCTCA 229
Qy 248 TCGCGCCCTCTCGGTCCTCTCGGCTGCTCACTGTTTCATGAGAAATGGGAGCTTGAGC 307
Db 230 TCGCGCCCTCTCGGTCCTCTCGGCTGCTCACTGTTTCATGAGAAATGGGAGCTTGAGC 289
Qy 308 CCGCGGCGGAGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 367
Db 290 CCGCGGCGGAGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 349
Qy 368 CGCACACCCCGCAGTGGCGCCATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 427
Db 350 CGCACACCCCGCAGTGGCGCCATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 409
Qy 428 GCGCGGAGCTGGGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 487
Db 410 GCGCGGAGCTGGGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 469
Qy 488 CTGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 547
Db 470 CTGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 529
Qy 548 GCTGGGAGAGCTGACGAGGAGATCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 607
Db 530 GCTGGGAGAGCTGACGAGGAGATCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 589
Qy 608 AGCTAAGGCTGCTGGGCGAGCGCCACTGTCATGTCCTACAGCCAGCCCGGCTCCCTCA 667
Db 590 AGCTAAGGCTGCTGGGCGAGCGCCACTGTCATGTCCTACAGCCAGCCCGGCTCCCTCA 649
Qy 668 ACCTCACTCTCCAGATATGCGAGGAGTGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 727
Db 650 ACCTCACTCTCCAGATATGCGAGGAGTGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 709
Qy 728 ACCTGCGCAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 787
Db 710 ACCTGCGCAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 769
Qy 788 AGCAGGAATACACAGCTTTGGGTTTGGCTGTGAGCGAGAAACCGCCTCGAGTTTCA 847
Db 770 AGCAGGAATACACAGCTTTGGCTTTGGCTGTGAGCGAGAAACCGCCTCGAGTTTCA 829
Qy 848 CTGCTGTGGCTTACCTATGAGCATGATACGAGGAGGAGTGTGCTGCTGCTGCTGCTGCTGCTG 907
Db 830 CTGCTGTGGCTTACCTATGAGCATGATACGAGGAGGAGTGTGCTGCTGCTGCTGCTGCTGCTG 889
Qy 908 CTGCTTTCACCCAGCCCGCAGAGACCCAGTCAGAT 945
Db 890 CTGCTTTCACCCAGCCCGCAGAGACCCAGTCAGAT 927
RESULT 4
AX480935
LOCUS AX480935 2681 bp DNA linear PAT 12-AUG-2002
DEFINITION Sequence 31 from Patent WO0246383.
ACCESSION AX480935
VERSION AX480935.1 GI:22217574
KEYWORDS Homo sapiens (human)
SOURCE

```
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1
REFERENCE
AUTHORS Yue,H., Azimzai,Y., Kallick,D.A., Baughn,M.R., Griffin,J.A.,
Swanekar,A., Lal,P.G., Wallia,N.K., Hafalia,A.J., Gandhi,A.R.,
Au-Young,J., Elliott,V.S., Ramkumar,J., Thangavelu,K., Lu,Y.,
Warren,B.A., Lu,D.A., Lee,E.A., Tribouley,C.M., Arvizu,C.,
Deleageane,A.M., Yao,M.G., Khan,F.A. and Sanjanwala,M.M.
TITLE Protein modification and maintenance molecules
JOURNAL Patent: WO 0246383-A 31 13-JUN-2002;
INCYTE Genomics, Inc. (US)
FEATURES
Location/Qualifiers
source
1..2681
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"
/notes="Incyte ID No: 2751509CB1"
ORIGIN
Query Match 65.0%; Score 716; DB 6; Length 2681;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 816; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 128 ACTGCGGGCCCTGAGCCCTCGGCCCGCATCGTGGGGGCTCAAAACGGCAGCGGGCA 187
Db 110 ACTGCGGGGCCCTGAGCCCTCGGCCCGCATCGTGGGGGCTCAAAACGGCAGCGGGCA 169
QY 188 CTGGGCTTGGGAAGTGAAGCTGACCATGAGGTGGCCACATCTGCGGGGCTCCCTCA 247
Db 170 CTTGGCCCTTGGCAAGTGAAGCTGACCATGAGGTGGCCACATCTGCGGGGCTCCCTCA 229
QY 248 TCGCCCTCTCTGGTCTCTCGCTGCTCACTGTTTCATGAGAAATGGAACTTGGAGC 307
Db 230 TCGCCCTCTCTGGTCTCTCGCTGCTCACTGTTTCATGAGAAATGGAACTTGGAGC 289
QY 308 CCGCGGCGGAGTGGTGGTACTGCTGGGCGTGCACTCCACAGGACGGGCCCTCGAGCGG 367
Db 290 CCGCGGCGGAGTGGTGGTACTGCTGGGCGTGCACTCCACAGGACGGGCCCTCGAGCGG 349
QY 368 CGCACACCGCGAGTGGCGGCATGTTGGTGGCGGCACACTACAGCAAGTGGAGCTGG 427
Db 350 CGCACACCGCGAGTGGCGGCATGTTGGTGGCGGCACACTACAGCAAGTGGAGCTGG 409
QY 428 GCGCGCACTGGCCCTGCTGCGCTGCGCTCACCCGCGAGCTGGGCCCGCCCTGTGGC 487
Db 410 GCGCGCACTGGCCCTGCTGCGCTGCGCTCACCCGCGAGCTGGGCCCGCCCTGTGGC 469
QY 488 CTGCTGCTGCTGCGCCGCGCTCAACCGCTTGGTGGACCGGACCGCTGCTGGGCCACCG 547
Db 470 CTGCTGCTGCTGCGCCGCGCTCAACCGCTTGGTGGACCGGACCGCTGCTGGGCCACCG 529
QY 548 GCTGGGAGAGCTCCAGAGGACATCTGCTCTCCCTGGTGTACTACAGAGTGG 607
Db 530 GCTGGGAGAGCTCCAGAGGACATCTGCTCTCCCTGGTGTACTACAGAGTGG 589
QY 608 AGCTAAGGCTGCTGGGAGGACCACTGTCAATGTCTCTACAGCAGCCCGCTCCCTTCA 667
Db 590 AGCTAAGGCTGCTGGGAGGACCACTGTCAATGTCTCTACAGCAGCCCGCTCCCTTCA 649
QY 668 ACCTCACTCTCCAGATATTCAGGAGATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 727
Db 650 ACCTCACTCTCCAGATATTCAGGAGATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 709
QY 728 ACACCTGCGAGGTGACTCTGGGGGGCCCTGCTGTGTGTGTGTGTGTGTGTGTGTGTGT 787
Db 710 ACACCTGCGAGGTGACTCTGGGGGGCCCTGCTGTGTGTGTGTGTGTGTGTGTGTGTGT 769
QY 788 AGGAGGAATCACAGGCTTTGGCTTGGCTGTGACGAGAGAAACCGCCCTGGAGTTTCA 847
Db 770 AGGAGGATACACAGGCTTTGGCTTGGCTGTGACGAGAGAAACCGCCCTGGAGTTTCA 829
QY 848 CTGCTGTGGCTTACCTATGAGGCATGGATACGGGAGCAGGTGATGGGTTTACAGGCTGG 907
Db 830 CTGCTGTGGCTTACCTATGAGGCATGGATACGGGAGCAGGTGATGGGTTTACAGGCTGG 889
QY 908 CTGCTGTGGCTTACCTATGAGGCATGGATACGGGAGCAGGTGATGGGTTTACAGGCTGG 945
Db 890 CTGCTGTGGCTTACCTATGAGGCATGGATACGGGAGCAGGTGATGGGTTTACAGGCTGG 927
RESULT 5
LOCUS BD127529 2810 bp DNA linear PAT 18-SEP-2002
DEFINITION Primer for synthesizing full-length cDNA and use thereof.
ACCESSION BD127529
VERSION BD127529.1 GI:23222474
KEYWORDS JP 2002017375-A/2960.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 2810)
Ota,T., Nishikawa,T., Isogai,T., Hayashi,K., Ishii,S., Kawai,Y.,
Wakamatsu,A., Sugiyama,T., Nagai,K., Kojima,S., Otsuki,T. and
Koga,H.
TITLE Primer for synthesizing full-length cDNA and use thereof
JOURNAL Patent: JP 2002017375-A 2960 22-JAN-2002;
COMMENT HELIX RESEARCH INSTITUTE
OS Homo sapiens (human)
PN JP 2002017375-A/2960
PD 22-JAN-2002
PF 07-JUN-2000 JP 200253172
PI TOSHIO OTA,TETSUO NISHIKAWA,TAKAO ISOGAI,KOJI HAYASHI,SHIZUKO
PI ISHII,
PI YURI KAWAI,AI WAKAMATSU,TOMOYASU SUGIYAMA,KEIICHI NAGAI, PI
SHINICHI KOJIMA,
PI TETSUUI OTSUKI,HISASHI KOGA
PC C12N15/09,C07K14/47,C07K16/18,C12N1/15,C12N1/19,C12N1/21,C12N5/ PC
10,
PC C12P21/02,C12Q1/68//C12P21/08,G06F17/30,C12N15/00,C12N5/00 CC
Primer for synthesizing full-length cDNA and use thereof FH Key
FEATURES
FT CDS Location/Qualifiers
(300)..(2597).
source
1..2810
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
ORIGIN
Query Match 57.3%; Score 631; DB 6; Length 2810;
Best Local Similarity 99.6%; Pred. No. 0;
Matches 781; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 162 GGGGGGCTCAAGCGCGCAGCGCGCACCTGGCTTGGCAAGTGAGCCTGCACCATGAGG 221
Db 176 GGGGGGCTCAAGCGCGCAGCGCGCACCTGGCTTGGCAAGTGAGCCTGCACCATGAGG 235
QY 222 TGGCCACATCTTGGGGGGCTCCCTCATCGCCCTCTCTGGGTCTCTCCGTGCTCACTG 281
Db 236 TGGCCACATCTTGGGGGGCTCCCTCATCGCCCTCTCTGGGTCTCTCCGTGCTCACTG 295
QY 282 TTTTCATGAGAAATGGAGCGTTGGAGCCCGGGCGGAGTGGTGGTACTGCTGGCGTGA 341
Db 296 TTTTCATGAGAAATGGAGCGTTGGAGCCCGGGCGGAGTGGTGGTACTGCTGGCGTGA 355
QY 342 CTCCAGGAGCGGCCCTCTGGACGGCGCGCACACCCGCGCAGTGGCGCCATCGTGTGCC 401
Db 356 CTCCAGGAGCGGCCCTCTGGACGGCGCGCACACCCGCGCAGTGGCGCCATCGTGTGCC 415
QY 402 GGCCTAACTACAGCAAGTGGAGCTGGGGCGCGGACCTTGGCCCTGCTGCGCTGCCCTCAC 461
Db 416 GGCCTAACTACAGCAAGTGGAGCTGGGGCGCGGACCTTGGCCCTGCTGCGCTGCCCTCAC 475
```


ORGANISM Homo sapiens
 Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
 1 (bases 1 to 670)
 Ota.T., Nishikawa.T., Isogai.T., Hayashi.K., Ishii.S., Kawai.Y.,
 Wakamatsu.A., Sugiyama.T., Nagai.K., Kojima.S., Otsuki.T. and
 Koga.H.
 Primer for synthesizing full-length cDNA and use thereof
 Patent: JP 2002017375-A 650 22-JAN-2002;
 HELIX RESEARCH INSTITUTE
 OS Homo sapiens (human)
 PN JP 2002017375-A/650
 PD 22-JAN-2002
 PF 07-JUL-2000 JP 2000253172
 PI TOSHIO OTA,TETSUO NISHIKAWA,TAKAO ISOGAI,KOJI HAYASHI,SHIZUKO
 PI ISHII,
 PI YURI KAWAI,AI WAKAMATSU,TOMOYASU SUGIYAMA,KEIICHI NAGAI, PI
 SHINICHI KOJIMA,
 PI TETSUJI OTSUKI,HISASHI KOGA
 PC
 C12N15/09,C07K14/47,C07K16/18,C12N1/15,C12N1/19,C12N1/21,C12N5/ PC
 10, C12P21/02,C12P1/68//C12P21/08,G06F17/30,C12N15/00,C12N5/00 CC
 PC C12P21/02,C12P1/68//C12P21/08,G06F17/30,C12N15/00,C12N5/00 CC
 Primer for synthesizing full-length cDNA and use thereof FH Key
 Location/Qualifiers
 FT source 1..670
 FT Location/Qualifiers
 /organism="Homo sapiens (human)"
 /mol_type="genomic DNA"
 /db_xref="taxon:9606"

FEATURES
 source
 1..670
 /organism="Homo sapiens"
 /mol_type="genomic DNA"
 /db_xref="taxon:9606"

ORIGIN
 Query Match 31.4%; Score 346; DB 6; Length 670;
 Best Local Similarity 99.7%; Pred. No. 6.3e-171;
 Matches 396; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 162 GGGGGGCTCAACCGCGAGCGCGGACCTGGCCCTGGCAAGTGGAGCTGACCATGGAGG 221
 176 GGGGGGCTCAACCGCGAGCGCGGACCTGGCCCTGGCAAGTGGAGCTGACCATGGAGG 235
 222 TGGCCACATCTGGGGGGCTCCCTCATCGCCCGCTCTGGGTCTCTCCGCTGCTCACTG 281
 236 TGGCCACATCTGGGGGGCTCCCTCATCGCCCGCTCTGGGTCTCTCCGCTGCTCACTG 295
 282 TTTTCATGCAATGGGAGCTGGAGCGCGCGAGTGGTTCGCTACTGCTGGCGGTGCA 341
 296 TTTTCATGCAATGGGAGCTGGAGCGCGCGAGTGGTTCGCTACTGCTGGCGGTGCA 355
 342 CTCCAGGACGGGCGCTTGGAGCGCGCGACACCGCGCGAGTGGCGGCTGCTGGTGC 401
 356 CTCCAGGACGGGCGCTTGGAGCGCGCGACACCGCGCGAGTGGCGGCTGCTGGTGC 415
 402 GGCAACTACAGCAAGTGGAGCTGGGGCGGAGCTGGCCCTGCTGGCCCTGGCTCACC 461
 416 GGCAACTACAGCAAGTGGAGCTGGGGCGGAGCTGGCCCTGCTGGCCCTGGCTCACC 475
 462 CGCAGCTGGGCGCGCGCTGGTGGCTGTGCTGCTGCCCGGCTCACACCGCTTCGT 521
 476 CGCAGCTGGGCGCGCGCTGGTGGCTGTGCTGCTGCCCGGCTCACACCGCTTCGT 535
 522 GCACGGACCGGCTGCTGGGCGGACCGGCTGGGGAGAC 558
 536 GCACGGACCGGCTGCTGGGCGGACCGGCTGGGGAGAC 572

RESULT 10
 BDI26485
 LOCUS BDI26485 670 bp DNA linear PAT 18-SEP-2002
 DEFINITION Primer for synthesizing full-length cDNA and use thereof.
 ACCESSION BDI26485
 VERSION BDI26485.1 GI:23221430
 KEYWORDS JP 2002017375-A/1916.
 SOURCE Homo sapiens (human)

ORGANISM Homo sapiens
 Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
 1 (bases 1 to 670)
 Ota.T., Nishikawa.T., Isogai.T., Hayashi.K., Ishii.S., Kawai.Y.,
 Wakamatsu.A., Sugiyama.T., Nagai.K., Kojima.S., Otsuki.T. and
 Koga.H.
 Primer for synthesizing full-length cDNA and use thereof
 Patent: JP 2002017375-A 1916 22-JAN-2002;
 HELIX RESEARCH INSTITUTE
 OS Homo sapiens (human)
 PN JP 2002017375-A/1916
 PD 22-JAN-2002
 PF 07-JUL-2000 JP 2000253172
 PI TOSHIO OTA,TETSUO NISHIKAWA,TAKAO ISOGAI,KOJI HAYASHI,SHIZUKO
 PI ISHII,
 PI YURI KAWAI,AI WAKAMATSU,TOMOYASU SUGIYAMA,KEIICHI NAGAI, PI
 SHINICHI KOJIMA,
 PI TETSUJI OTSUKI,HISASHI KOGA
 PC
 C12N15/09,C07K14/47,C07K16/18,C12N1/15,C12N1/19,C12N1/21,C12N5/ PC
 10, C12P21/02,C12P1/68//C12P21/08,G06F17/30,C12N15/00,C12N5/00 CC
 PC C12P21/02,C12P1/68//C12P21/08,G06F17/30,C12N15/00,C12N5/00 CC
 Primer for synthesizing full-length cDNA and use thereof FH Key
 Location/Qualifiers
 FT source 1..670
 FT Location/Qualifiers
 /organism="Homo sapiens (human)"
 /mol_type="genomic DNA"
 /db_xref="taxon:9606"

FEATURES
 source
 1..670
 /organism="Homo sapiens"
 /mol_type="genomic DNA"
 /db_xref="taxon:9606"

ORIGIN
 Query Match 31.4%; Score 346; DB 6; Length 670;
 Best Local Similarity 99.7%; Pred. No. 6.3e-171;
 Matches 396; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 162 GGGGGGCTCAACCGCGAGCGCGGACCTGGCCCTGGCAAGTGGAGCTGACCATGGAGG 221
 176 GGGGGGCTCAACCGCGAGCGCGGACCTGGCCCTGGCAAGTGGAGCTGACCATGGAGG 235
 222 TGGCCACATCTGGGGGGCTCCCTCATCGCCCGCTCTGGGTCTCTCCGCTGCTCACTG 281
 236 TGGCCACATCTGGGGGGCTCCCTCATCGCCCGCTCTGGGTCTCTCCGCTGCTCACTG 295
 282 TTTTCATGCAATGGGAGCTGGAGCGCGCGAGTGGTTCGCTACTGCTGGCGGTGCA 341
 296 TTTTCATGCAATGGGAGCTGGAGCGCGCGAGTGGTTCGCTACTGCTGGCGGTGCA 355
 342 CTCCAGGACGGGCGCTTGGAGCGCGCGACACCGCGCGAGTGGCGGCTGCTGGTGC 401
 356 CTCCAGGACGGGCGCTTGGAGCGCGCGACACCGCGCGAGTGGCGGCTGCTGGTGC 415
 402 GGCAACTACAGCAAGTGGAGCTGGGGCGGAGCTGGCCCTGCTGGCCCTGGCTCACC 461
 416 GGCAACTACAGCAAGTGGAGCTGGGGCGGAGCTGGCCCTGCTGGCCCTGGCTCACC 475
 462 CGCAGCTGGGCGCGCGCTGGTGGCTGTGCTGCTGCCCGGCTCACACCGCTTCGT 521
 476 CGCAGCTGGGCGCGCGCTGGTGGCTGTGCTGCTGCCCGGCTCACACCGCTTCGT 535
 522 GCACGGACCGGCTGCTGGGCGGACCGGCTGGGGAGAC 558
 536 GCACGGACCGGCTGCTGGGCGGACCGGCTGGGGAGAC 572

RESULT 11
 AC009088/c
 LOCUS AC009088 127769 bp DNA linear PRI 29-MAR-2003
 DEFINITION Homo sapiens chromosome 16 clone RP11-388M20, complete sequence.
 ACCESSION AC009088
 VERSION AC009088.9 GI:29366934

```
KEYWORDS      HTG.
SOURCE        Homo sapiens (human)
ORGANISM      Homo sapiens
              Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
              Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE     1 (bases 1 to 127769)
AUTHORS      DOE Joint Genome Institute, Stanford Human Genome Center and Los
              Alamos National Laboratory.
TITLE        Direct Submission
JOURNAL      Unpublished
REFERENCE     2 (bases 1 to 127769)
AUTHORS      DOE Joint Genome Institute.
TITLE        Direct Submission
JOURNAL      Submitted (03-APR-2002) Production Sequencing Facility, DOE Joint
              Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
REFERENCE     3 (bases 1 to 127769)
AUTHORS      DOE Joint Genome Institute.
TITLE        Direct Submission
JOURNAL      Submitted (03-AUG-1999) Production Sequencing Facility, DOE Joint
              Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
REFERENCE     4 (bases 1 to 127769)
AUTHORS      DOE Joint Genome Institute, Stanford Human Genome Center and Los
              Alamos National Laboratory.
TITLE        Direct Submission
JOURNAL      Submitted (18-MAR-2003) DOE Joint Genome Institute, 2800 Mitchell
              Drive, Walnut Creek, CA 94598, USA
REFERENCE     5 (bases 1 to 127769)
AUTHORS      DOE Joint Genome Institute, Stanford Human Genome Center and Los
              Alamos National Laboratory.
TITLE        Direct Submission
JOURNAL      Submitted (29-MAR-2003) DOE Joint Genome Institute, 2800 Mitchell
              Drive, Walnut Creek, CA 94598, USA
COMMENT      On Mar 29, 2003 this sequence version replaced gi:29029216.
              Draft Sequence Produced by DOE Joint Genome Institute
              www.jgi.doe.gov
              Location/Qualifiers
                1..127769
                  /organism="Homo sapiens"
                  /mol_type="genomic DNA"
                  /db_xref="taxon:9606"
                  /chromosome="16"
                  /clone="RF11-388M20"

ORIGIN
Query Match      24.5%; Score 270; DB 9; Length 127769;
Best Local Similarity 100.0%; Pred. No. 5.8e-131;
Matches 270; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 302 TGGAGCCCGCGCGAGTGTGCGTACTGTCTGGCGCTGCTACTCCAGGACGGCCCTGG 361
      |||||
Db 31579 TGGAGCCCGCGCGCGAGTGTGCGTACTGTCTGGCGCTGCTACTCCAGGACGGCCCTGG 31520

QY 362 ACGCGCGCGACACCCGCGAGTGGCGCCGCTGTGTGTGCGCGCCCACTACAGCCAAGTGG 421
      |||||
Db 31519 ACGCGCGCGACACCCGCGAGTGGCGCCGCTGTGTGTGCGCGCCCACTACAGCCAAGTGG 31460

QY 422 AGCTGGCGCGGACTGGCCCTGTGCGCTGGCTCACCAGCCAGCTGGCGCCCGCCGCG 481
      |||||
Db 31459 AGCTGGCGCGGACTGGCCCTGTGCGCTGGCTCACCAGCCAGCTGGCGCCCGCCGCG 31400

QY 482 TGTGCTGTGTGCTGTGCTGCGCGCGCTCACACCGCTTGTGTGTCAGGACCGCCCTGTGGG 541
      |||||
Db 31399 TGTGCTGTGTGCTGTGCTGCGCGCGCTCACACCGCTTGTGTGTCAGGACCGCCCTGTGGG 31340

QY 542 CCACCGGCTGGGAGACGTCCAGGAGGAG 571
```

Db 31339 CCACCGGCTGGGAGACGTCCAGGAGGAG 31310

RESULT 12
AC135044

LOCUS
DEFINITION

AC135044.1 GI:23505535
HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ACTIVEFIN.

ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

AC135044.1 GI:23505535
HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ACTIVEFIN.

AC135044.1 GI:23505535
HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ACTIVEFIN.

AC135044.1 GI:23505535
HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ACTIVEFIN.

AC135044.1 GI:23505535
HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ACTIVEFIN.

AC135044.1 GI:23505535
HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ACTIVEFIN.

AC135044.1 GI:23505535
HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ACTIVEFIN.

AC135044.1 GI:23505535
HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ACTIVEFIN.

AC135044.1 GI:23505535
HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ACTIVEFIN.

AC135044.1 GI:23505535
HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ACTIVEFIN.

AC135044.1 GI:23505535
HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ACTIVEFIN.

AC135044.1 GI:23505535
HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ACTIVEFIN.

AC135044.1 GI:23505535
HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ACTIVEFIN.

AC135044.1 GI:23505535
HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ACTIVEFIN.

AC135044.1 GI:23505535
HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ACTIVEFIN.

AC135044.1 GI:23505535
HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ACTIVEFIN.

AC135044.1 GI:23505535
HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ACTIVEFIN.

AC135044.1 GI:23505535
HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ACTIVEFIN.

AC135044.1 GI:23505535
HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ACTIVEFIN.

AC135044.1 GI:23505535
HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ACTIVEFIN.

AC135044.1 GI:23505535
HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ACTIVEFIN.

AC135044.1 GI:23505535
HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ACTIVEFIN.

AC135044.1 GI:23505535
HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ACTIVEFIN.

AC135044.1 GI:23505535
HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ACTIVEFIN.

AC135044.1 GI:23505535
HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ACTIVEFIN.

AC135044.1 GI:23505535
HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ACTIVEFIN.

AC135044.1 GI:23505535
HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ACTIVEFIN.

AC135044.1 GI:23505535
HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ACTIVEFIN.

AC135044.1 GI:23505535
HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ACTIVEFIN.

AC135044.1 GI:23505535
HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ACTIVEFIN.

AC135044.1 GI:23505535
HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ACTIVEFIN.

AC135044.1 GI:23505535
HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ACTIVEFIN.

AC135044.1 GI:23505535
HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ACTIVEFIN.

AC135044.1 GI:23505535
HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ACTIVEFIN.

```
* 32820 32919: gap of unknown length
* 32920 39003: contig of 6084 bp in length
* 39004 39103: gap of unknown length
* 39104 48370: contig of 9267 bp in length
* 48371 48470: gap of unknown length
* 48471 59721: contig of 11251 bp in length
* 59722 59821: gap of unknown length
* 59822 72304: contig of 12483 bp in length
* 72305 72404: gap of unknown length
* 72405 85349: contig of 12945 bp in length
* 85350 85449: gap of unknown length
* 85450 101872: contig of 16423 bp in length
* 101873 117076: gap of unknown length
* 117077 117176: gap of unknown length
* 117177 138973: contig of 21797 bp in length
* 138974 139074: gap of unknown length
* 139074 150337: contig of 11263 bp in length
* 150337 150437: gap of unknown length
* 150437 195476: contig of 45040 bp in length.
FEATURES             Location/Qualifiers
     source            1..195476
     organism="Homo sapiens"
     /mol_type="genomic DNA"
     /db_xref="taxon:9606"
     /chromosome="16"
     /clone="CTD-2551B20"
     /clone_lib="CalTech human BAC library D"

ORIGIN
Query Match          24.5%; Score 270; DB 2; Length 195476;
Best Local Similarity 100.0%; Pred. No. 5.6e-131;
Matches 270; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 302 TGGAGCCGCGCGCGAGTGTGCGTACTGCTGGGCGTGCATCCAGAGACGGGCCCTGG 361
Db 27417 TGGAGCCGCGCGCGAGTGTGCGTACTGCTGGGCGTGCATCCAGAGACGGGCCCTGG 27476

Qy 362 ACGGCGCGCACACCGCGCATGCGCGCATGCGTGGCGCGCACTACAGCAAGTGG 421
Db 27477 ACGGCGCGCACACCGCGCATGCGCGCATGCGTGGCGCGCACTACAGCAAGTGG 27536

Qy 422 AGCTGGGCGCGCGACCTGCGCCCTGCTGCGCCCTGCTACCGCGCGACGCTGGCGCCCGCG 481
Db 27537 AGCTGGGCGCGCGACCTGCGCCCTGCTGCGCCCTGCTACCGCGCGACGCTGGCGCCCGCG 27596

Qy 482 TGTGGCCTGTGCTGCTGCGCGCGCTCACACCGCTTGTGTCAGCGGACCGCGCTGCTGG 541
Db 27597 TGTGGCCTGTGCTGCTGCGCGCGCTCACACCGCTTGTGTCAGCGGACCGCGCTGCTGG 27656

Qy 542 CCACCGCTGGGAGACGCTCCAGGAGCGAG 571
Db 27657 CCACCGCTGGGAGACGCTCCAGGAGCGAG 27686

RESULT 13
AC093520/c          168064 bp DNA linear PRI 19-MAR-2003
LOCUS              Homo sapiens chromosome 16 clone RP11-120K18, complete sequence.
DEFINITION          AC093520
ACCESSION            AC093520
VERSION              AC093520.4 GI:29124069
KEYWORDS             HTG.
SOURCE               Homo sapiens (human)
ORGANISM             Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE            1 (bases 1 to 168064)
AUTHORS              DOE Joint Genome Institute, Stanford Human Genome Center and Los
                     Alamos National Laboratory.
TITLE                Direct Submission
JOURNAL              Unpublished
REFERENCE            2 (bases 1 to 168064)
AUTHORS              DOE Joint Genome Institute.
```

```
Direct Submission
Submitted (31-AUG-2001) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
3 (bases 1 to 168064)
DOE Joint Genome Institute.
Direct Submission
Submitted (13-JUN-2002) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
4 (bases 1 to 168064)
DOE Joint Genome Institute.
Direct Submission
Submitted (07-JUL-2002) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
5 (bases 1 to 168064)
DOE Joint Genome Institute, Stanford Human Genome Center and Los
Alamos National Laboratory.
Direct Submission
Submitted (19-MAR-2003) DOE Joint Genome Institute, 2800 Mitchell
Drive, Walnut Creek, CA 94598, USA
On Mar 19, 2003 this sequence version replaced gi:21702833.
Draft Sequence Produced by DOE Joint Genome Institute
www.jgi.doe.gov
Finishing Completed at Stanford Human Genome Center and Los Alamos
National Laboratory
www.sbgc.stanford.edu
Quality: Phrap Quality >=40 100% of Sequence;
Estimated Total Number of Errors is 0.
FEATURES             Location/Qualifiers
     source            1..168064
     organism="Homo sapiens"
     /mol_type="genomic DNA"
     /db_xref="taxon:9606"
     /chromosome="16"
     /clone="RP11-120K18"

ORIGIN
Query Match          7.9%; Score 87; DB 9; Length 168064;
Best Local Similarity 100.0%; Pred. No. 2.5e-34;
Matches 87; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 952 CATCAACGGCATCTCGGATCTCGCCAGATCCCTTTGAGGCCCTGTCCCATATATCA 1011
Db 141592 CATCAACGGCATCTCGGATCTCGCCAGATCCCTTTGAGGCCCTGTCCCATATATCA 141533

Qy 1012 GTAGGAGTCTCACTGGGACCAAAAGC 1038
Db 141532 GTAGGAGTCTCACTGGGACCAAAAGC 141506

RESULT 14
AX098193            1796 bp DNA linear PAT 30-MAR-2001
LOCUS              Sequence 105 from Patent WO0118542.
DEFINITION          AX098193
ACCESSION            AX098193
VERSION              AX098193.1 GI:13515276
KEYWORDS             Homo sapiens (human)
SOURCE               Homo sapiens
ORGANISM             Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE            1
AUTHORS              Lee, J., Thompson, P. and Lillie, J.
TITLE                Identification, assessment, prevention, and therapy of ovarian
                     cancer
JOURNAL              Patent: WO 0118542-A 105 15-MAR-2001;
                     Millennium Predictive Medicine, Inc. (US)
FEATURES             Location/Qualifiers
     source            1..1796
     organism="Homo sapiens"
     /mol_type="unassigned DNA"
     /db_xref="taxon:9606"

ORIGIN
```

Query Match 6.1%; Score 67; DB 6; Length 1796;
 Best Local Similarity 100.0%; Pred. No. 1.3e-23;
 Matches 67; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GGGCCCTGTCTCTGGCCATGCCCGAGAGGGGCTCTGGGGCTGGGCGACCTGGGGGCT 60
 |||||
 DB 134 GGGCCCTGTCTCTGGCCATGCCCGAGAGGGGCTCTGGGGCTGGGCGACCTGGGGGCT 193
 |||||

QY 61 GTGGCCA 67
 |||||
 DB 194 GTGGCCA 200
 |||||

RESULT 15
 BC001462
 LOCUS BC001462 1800 bp mRNA linear PRI 04-OCT-2003
 DEFINITION Homo sapiens protease, serine, 8 (proctasin), mRNA (cDNA clone
 MGC:2133 IMAGE:3138532), complete cds.
 ACCESSION BC001462
 VERSION BC001462.2 GI:33876244
 KEYWORDS MGC.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (bases 1 to 1800)
 AUTHORS Klausner, R.D., Collins, F.S., Wagner, L.H., Derge, J.G.,
 Altschul, S.F., Zeeberg, B., Buetow, K.H., Schaefer, C.F., Bhat, N.K.,
 Hopkins, R.F., Jordan, H., Moore, T., Max, S.I., Wang, J.F., Haieh, F.,
 Diatchenko, L., Marusina, K., Farmer, A.A., Rubin, G.M., Hong, L.,
 Stapleton, M., Soares, M.B., Bonaldo, M.F., Casavant, T.L.,
 Scheetz, T.E., Brownstein, M.J., Usdin, T.B., Toehiyuki, S.,
 Carninci, P., Prange, C., Raha, S.S., Loquellano, N.A., Peters, G.J.,
 Abramson, R.D., Mullighy, S.J., Bosak, S.A., McEwan, P.J.,
 McKernan, K.J., Malek, J.A., Gunaratne, P.H., Richards, S.,
 Worley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Hulyk, S.W.,
 Villalón, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A.,
 Fahey, J., Helton, E., Kettman, M., Madan, A., Rodriguez, S.,
 Sanchez, A., Whiting, M., Madan, A., Young, A.C., Shevchenko, Y.,
 Bouffard, G.G., Blakesley, R.W., Touchman, J.W., Green, E.D.,
 Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M.,
 Butlerfield, Y.S., Krzywinski, M.I., Skalska, U., Smalley, D.E.,
 Schneringer, A., Schein, J.E., Jones, S.J., Skalska, U., Smalley, D.E.,
 human and mouse cDNA analyses of more than 15,000 full-length
 human and mouse cDNA sequences
 Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
 22388257
 12477932
 REFERENCE 2 (bases 1 to 1800)
 AUTHORS Strausberg, R.
 TITLE Direct Submission
 JOURNAL Submitted (12-DEC-2000) National Institutes of Health, Mammalian
 Gene Collection (MGC), Cancer Genomics Office, National Cancer
 Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
 USA
 NIH-MGC Project URL: <http://img.nci.nih.gov>
 On Aug 19, 2003 this sequence version replaced gi:12655206.
 Contact: MGC help desk
 Email: cgapbs-r@mail.nih.gov
 Tissue Procurement: ATCC
 cDNA Library Preparation: Rubin Laboratory
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: National Institutes of Health Intramural
 Sequencing Center (NISC),
 Gaithersburg, Maryland;
 Web site: <http://www.nisc.nih.gov/>
 Contact: nisc.mgc@nih.gov
 Akhtar, N., Ayele, K., Beckstrom-Sternberg, S.M., Benjamin, B.,
 Blakesley, R.W., Bouffard, G.G., Breen, K., Brinkley, C., Brooks, S.,
 Dietrich, N.L., Granite, S., Guan, X., Gupta, J., Haghighi, P.,
 Hansen, N., Ho, S.-L., Karlins, E., Kwong, P., Lario, P., Legaapi, R.,
 Maduro, Q.L., Masello, C., Maskeri, B., Mastrian, S.D., McCloskey, J.C.,

McDowell, J., Pearson, R., Stantripop, S., Thomas, P.J., Touchman, J.W.,
 Tsugeon, C., Vogt, J.L., Walker, M.A., Wetherby, K.D., Wiggins, L.,
 Young, A., Zhang, L.-H. and Green, E.D.

Clone distribution: MGC clone distribution information can be found
 through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
 Series: IRAL Plate: 4 Row: J Column: 3
 This clone was selected for full length sequencing because it
 passed the following selection criteria: matched mRNA gi: 21536453.

FEATURES
 Location/Qualifiers
 1..1800
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="MGC:2133 IMAGE:3138532"
 /tissue_type="Placenta, Choriocarcinoma"
 /clone_lib="NIH_MGC_21"
 /lab_host="DH10B-R"
 /note="Vector: pOT7"
 1..1800
 /gene="PRSS8"
 /note="synonym: PROSTASIN"
 /db_xref="LocusID:5652"
 176..1207
 /db_xref="MIM:600823"
 176..1207
 /codon_start=1
 /product="proctasin preproprotein"
 /protein_id="AAH01462.1"
 /db_xref="GI:12655207"
 /db_xref="LocusID:5652"
 ITGSSAVAGQWQVTSITVEGVVCGSGSLVSEQWLSAAHCFSEHKEAYEVKGA
 HOLDVSADAKVTLKDIIPHSYLOBSQDIALLOLRPTIFSRIRPTICLPAANA
 SFNGLHCTVTGHWAPSLSLTPELQLEVLISRTNCLYNIDAKPEPHFVQ
 EDWVCAGYVGGDKACGDSGSLSCVEGLWLTGLVSWGDAGGANNRPQVYTLAS
 YASWIOGKVTQLQPRVVVPTQESQPSDNLGSHLAFSSAPAQGLRILFLPLGLALG
 LLSFWLSEH"
 305..1018
 /note="Tryp_SPC: Region: Trypsin-like serine protease"
 /db_xref="CDD:smart00020"

misc_feature
 305..1018
 /note="Tryp_SPC: Region: Trypsin-like serine protease"

ORIGIN
 Query Match 6.1%; Score 67; DB 9; Length 1800;
 Best Local Similarity 100.0%; Pred. No. 1.3e-23;
 Matches 67; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GGGCCCTGTCTCTGGCCATGCCCGAGAGGGGCTCTCTGGGGCTGGGCGACCTGGGGGCT 60
 |||||
 DB 158 GGGCCCTGTCTCTGGCCATGCCCGAGAGGGGCTCTCTGGGGCTGGGCGACCTGGGGGCT 217
 |||||

QY 61 GTGGCCA 67
 |||||
 DB 218 GTGGCCA 224
 |||||

Search completed: February 25, 2004, 13:48:54
 Job time : 4330 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 25, 2004, 11:17:50 ; Search time 475 Seconds
(without alignments)

9855.820 Million cell updates/sec

Title: US-10-037-417-45

Perfect score: 1102

Sequence: 1 gggccctgtcctgggcat.....gggtcttgatggggctcc 1102

Scoring table: OLIGO_NUC

Gapop 60.0 , Gapext 60.0

Searched: 3373863 seqs, 2124099041 residues

Word size : 0

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

N_Geneseq_29Jan04.*

1: Geneseqn1980s.*

2: Geneseqn1990s.*

3: Geneseqn2000s.*

4: Geneseqn2001as.*

5: Geneseqn2001bs.*

6: Geneseqn2002s.*

7: Geneseqn2003as.*

8: Geneseqn2003bs.*

9: Geneseqn2003cs.*

10: Geneseqn2004s.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1102	100.0	1102	6 ABQ93901	Abq93901 Human pro
2	1102	100.0	1102	6 ABQ93902	Abq93902 Human pro
3	716	65.0	2457	6 ABK31795	Abk31795 DNA encod
4	716	65.0	2681	6 ABQ75956	Abq75956 Human PMM
5	631	57.3	2810	4 AAK94500	AAK94500 Human ful
6	486	44.1	537	6 ABK13565	Abk13565 Human pro
7	398	36.1	597	5 AAS69040	AAS69040 DNA encod
8	384	34.8	456	6 ABK13566	Abk13566 Human pro
9	346	31.4	670	4 AAK92190	AAK92190 Human CDN
10	346	31.4	670	4 AAK93456	AAK93456 Human CDN
11	135	12.3	768	4 AAH78206	AAH78206 Nucleotid
12	135	12.3	2298	5 AAS71079	AAS71079 DNA encod
13	135	12.3	1733	7 ABT31936	ABT31936 Human bre
14	67	6.1	1796	5 AAF98698	AAF98698 Human ova
15	67	6.1	1835	5 AAF98720	AAF98720 Human lat
16	67	6.1	3382	6 ABZ35336	ABZ35336 Human gen
17	53	4.8	1726	6 ABN85392	ABN85392 Human NOV
18	53	4.8	1834	6 ABM76501	ABM76501 cDNA enco
19	53	4.8	1834	6 ABT67949	ABT67949 Ovary can
20	53	4.8	1834	6 ABT68512	ABT68512 Kidney ca
21	53	4.8	1834	6 ABT68248	ABT68248 Kidney ca
22	53	4.8	1834	6 ABK12241	ABK12241 cDNA enco
23	53	4.8	1834	6 ABN95716	ABN95716 Gene #221

ALIGNMENTS

RESULT 1

ABQ93901

ID ABQ93901 standard; DNA; 1102 BP.

XX AC

XX AC

XX AC

XX AC

DT 01-NOV-2002 (first entry)

DE Human prostatic precursor-like NOV14a DNA, SEQ ID NO:43.

XX XX

XX XX

XX XX

XX XX

XX XX

XX XX

XX XX

XX XX

XX XX

XX XX

XX XX

XX XX

XX XX

XX XX

XX XX

XX XX

XX XX

XX XX

XX XX

XX XX

XX XX

XX XX

XX XX

XX XX

XX XX

XX XX

XX XX

XX XX

XX XX

XX XX

XX XX

XX XX

XX XX

XX XX

XX XX

XX XX

XX XX

XX XX

XX XX

XX XX

XX XX

XX XX

XX XX

XX XX

Aai29377 Colon tum
Abz33563 Human col
Aba21236 Human ner
Abn85393 Human NOV
Aac77814 Human can
Aas26531 Human CDN
Abx73972 Human nov
Aas26109 Human CDN
Abx73450 Human nov
Abk13570 Prostatin
Abq93931 Human NOV
Abs52114 Human bet
Abl29631 Drosophil
Abl29630 Drosophil
Abl18082 Drosophil
Aad27724 Human tra
Abk30252 Human G-p
Abq83347 Human KLK
Aaa95896 Human KLK
Abk31801 DNA encod
Abk30259 Human G-p
Abk30258 Human G-p

ALIGNMENTS

RESULT 1

ABQ93901

ID ABQ93901 standard; DNA; 1102 BP.

XX AC

XX AC

XX AC

XX AC

DT 01-NOV-2002 (first entry)

DE Human prostatic precursor-like NOV14a DNA, SEQ ID NO:43.

XX XX

XX XX

XX XX

XX XX

XX XX

XX XX

XX XX

XX XX

XX XX

XX XX

XX XX

XX XX

XX XX

XX XX

XX XX

XX XX

XX XX

XX XX

XX XX

XX XX

XX XX

XX XX

XX XX

XX XX

XX XX

XX XX

XX XX

XX XX

XX XX

XX XX

XX XX

XX XX

XX XX

XX XX

XX XX

XX XX

07-JAN-2002; 2002WO-US000375.

05-JAN-2001; 2001US-0260018P.

08-JAN-2001; 2001US-0260360P.

28-FEB-2001; 2001US-0272411P.

/product= "NOV14a"

/note= "No stop codon given in the specification"

WO200253742-A2.

11-JUL-2002.

PR 02-MAR-2001; 2001US-0272817P.
PR 05-JUL-2001; 2001US-0302318P.
PR 12-JUL-2001; 2001US-0305060P.
PR 10-SEP-2001; 2001US-0318405P.
PR 12-SEP-2001; 2001US-0318700P.
PR 04-JAN-2002; 2002US-00037417.
XX
PA (CURA-) CURAGEN CORP.
XX
XX Kekuda R, Alsobrook JP, Tchernev VT, Liu X, Spytek KA;
PI Patturajan M, Grosse WM, Lepley DM, Burgess CE, Vernet CM, Li L;
PI Gorman L, Edinger S, Sciore P, Ellerman K, Malyankar U;
PI Rothenberg M, Stone D, Boldog F, Guo X, Shency S, Anderson D;
PI Padigaru M, Taupier RJ, Miller CE, Eisen A;
XX
XX WPI; 2002-583619/62.
DR P-PSDB; ABB09523.
XX
XX Novel polypeptides and nucleic acids homologous to transmembrane
PT receptor, thymosin, neuromodulin-like family of proteins for diagnosing,
PT treating cancer, atherosclerosis, neurological, skin and autoimmune
PT disorders.
XX
XX Claim 9a; Page 142; 323pp; English.
XX
XX The invention relates to 24 novel human proteins designated NOV1-NOV14
CC (ABB09501-ABB09524), collectively referred to as NOVX proteins, and
CC nucleic acids encoding them (ABQ93879-ABQ93902). NOVX proteins and
CC nucleotides are useful in the treatment, diagnosis or prevention of NOVX-
CC associated disorders or in the manufacture of a medicament for treating
CC such disorders, with specific applications described for each of the 24
CC NOVX proteins, based on their homology to known proteins. Various
CC disorders are associated with NOVX proteins including neurological
CC disorders (e.g., Alzheimer's, Huntington's and Parkinson's diseases),
CC pain, behavioural disorders, addiction, tuberculous sclerosis, cancers
CC (e.g., colorectal cancer, leukaemia and osteosarcoma), immune disorders
CC (e.g., allergies and autoimmune diseases), myasthenia gravis, asthma,
CC various forms of arthritis, diabetes, thyroiditis, cardiovascular disease
CC (e.g., hypertension), reproductive disorders, endometriosis,
CC incontinence, psoriasis, scleroderma, alopecia, ulcers, pancreatitis,
CC cirrhosis, glomerular endotheliosis, polycystic kidney disease, endocrine
CC disorders, obesity, bacterial infections and particularly cardiomyopathy,
CC atherosclerosis, cell signal processing-related disorders and disorders
CC of metabolic pathway regulation. NOVX nucleic acids and polypeptides may
CC be used to identify cellular receptors or downstream effectors which
CC binds to a NOVX protein, and are also useful as targets for the
CC identification of small molecules that modulate or inhibit processes such
CC as neurogenesis, cell differentiation, cell motility, cellular
CC proliferation, haematopoiesis, wound healing and angiogenesis. NOVX
CC nucleic acid sequences can be used to identify a cell or tissue type and
CC are useful as a source of primers or probes for forensic biology and for
CC identifying and cloning NOVX homologues in other cell types. Cells
CC comprising NOVX nucleic acids are useful for producing non-human
CC transgenic animals which are useful for studying the function and
CC activity of NOVX proteins and for identifying and evaluating modulators
CC of NOVX activity. The present sequence represents DNA encoding the
CC prostatic precursor-like protein NOV14a. The gene encoding NOV14a is
CC located on chromosome 16
XX
SQ Sequence 1102 BP; 162 A; 371 C; 353 G; 216 T; 0 U; 0 Other;
Query Match 100.0%; Score 1102; DB 6; Length 1102;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1102; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GGGGCCCTTCTCGGCCCATGGCCCAAGAGGGGGTCTCGGGCCCTGGGAGCTGGGGCT 60
DB 1 GGGGCCCTTCTCGGCCCATGGCCCAAGAGGGGGTCTCGGGCCCTGGGAGCTGGGGCT 60
QY 61 GTGGCCAAATTCGACTCATCTACCTTACCGGTTGGTCCGCGACCCCGCTAGGGGC 120
DB 61 GTGGCCAAATTCGACTCATCTACCTTACCGGTTGGTCCGCGACCCCGCTAGGGGC 120

QY 121 CCCCCTACTGCGGGCGCCCTGAGCCCTCGGCCCGCATCTGTTGGGGGGCTCAAAACGCGCAG 180
DB 121 CCCCCTACTGCGGGCGCCCTGAGCCCTCGGCCCGCATCTGTTGGGGGGCTCAAAACGCGCAG 180
QY 181 CCGGGCACTGGCCCTTGGCAAGTGGAGCTGACCATGAGAGTGGCCACATCTGCGGGGGC 240
DB 181 CCGGGCACTGGCCCTTGGCAAGTGGAGCTGACCATGAGAGTGGCCACATCTGCGGGGGC 240
QY 241 TCCCTCATCGCCCTCTCTGGGTCTCTCCGCTGCTCACTGTTTATGACGAATGGGAGC 300
DB 241 TCCCTCATCGCCCTCTCTGGGTCTCTCCGCTGCTCACTGTTTATGACGAATGGGAGC 300
QY 301 TTGGAGCCCGCGCGAGTGGTGTGCTGCTGGGGGTGACATCCAGAGCGGGCCCTG 360
DB 301 TTGGAGCCCGCGCGAGTGGTGTGCTGCTGGGGGTGACATCCAGAGCGGGCCCTG 360
QY 361 GACGGCGGCACACCGCGCGAGTGGTGTGCTGGGGGTGACATCCAGAGCGGGCCCTG 420
DB 361 GACGGCGGCACACCGCGCGAGTGGTGTGCTGGGGGTGACATCCAGAGCGGGCCCTG 420
QY 421 GAGCTGGGGCGCGAGCTGGCCCTGCTGGCGCTGACCCGCGCGAGCTGGGGCCCTG 480
DB 421 GAGCTGGGGCGCGAGCTGGCCCTGCTGGCGCTGACCCGCGCGAGCTGGGGCCCTG 480
QY 481 GTGTGGCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 540
DB 481 GTGTGGCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 540
QY 541 GGCACCGCTGGGGAGAGCTGCCAGGAGGAGATCTCTGCTCTCCCTGGGTGCTACAG 600
DB 541 GGCACCGCTGGGGAGAGCTGCCAGGAGGAGATCTCTGCTCTCCCTGGGTGCTACAG 600
QY 601 GAAGTGGAGCTAAGGCTGCTGGGGCGAGGCGACCTGTCATGCTCTACAGCGCCGCT 660
DB 601 GAAGTGGAGCTAAGGCTGCTGGGGCGAGGCGACCTGTCATGCTCTACAGCGCCGCT 660
QY 661 CCCTTCAACT 720
DB 661 CCCTTCAACT 720
QY 721 CGCAGGAGACACTGCGCAGGCTGACTCTGGGGGGCCCTGCTGCTGAGGAAGCGGGCGC 780
DB 721 CGCAGGAGACACTGCGCAGGCTGACTCTGGGGGGCCCTGCTGCTGAGGAAGCGGGCGC 780
QY 781 TGGTTCCAGCAGGAATCACAGCTTTGGGTGCTGGAGCGAGAGAAACCGCCCTGGA 840
DB 781 TGGTTCCAGCAGGAATCACAGCTTTGGGTGCTGGAGCGAGAGAAACCGCCCTGGA 840
QY 841 GTTTTCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 900
DB 841 GTTTTCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 900
QY 901 CTTGGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 960
DB 901 CTTGGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 960
QY 961 GCATTCCTGATTTCTGCGAGATCTTTTGGGGCCCTTGTCCCATATATCATAGTAGGAGTC 1020
DB 961 GCATTCCTGATTTCTGCGAGATCTTTTGGGGCCCTTGTCCCATATATCATAGTAGGAGTC 1020
QY 1021 TCAACTGGAGCAAAAGCTTGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1080
DB 1021 TCAACTGGAGCAAAAGCTTGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1080
QY 1081 TGGGGGTTCTGATGGGGCTCC 1102
DB 1081 TGGGGGTTCTGATGGGGCTCC 1102
RESULT 2
ID ABQ93902 standard; DNA; 1102 BP.
XX

601 GAAGTGGAGTAAAGTCTGCTGGGAGGAGCCACTGTCAATGTCTCTACAGCCAGCCCGGT 660
 PT
 601 GAAGTGGAGTAAAGTCTGCTGGGAGGAGCCACTGTCAATGTCTCTACAGCCAGCCCGGT 660
 Db
 661 CCCTTTCAACCTCACTCTCCAGATATTGCGAGGATGCTGTGCTGGCTTACCCAGAGGC 720
 QY
 661 CCCTTTCAACCTCACTCTCCAGATATTGCGAGGATGCTGTGCTGGCTTACCCAGAGGC 720
 Db
 721 CGCAGGAGACCTCCAGAGTCACTCTGGGGGCGCTGTGTGAGGAAGGCGCGC 780
 QY
 721 CGCAGGAGACCTCCAGAGTCACTCTGGGGGCGCTGTGTGAGGAAGGCGCGC 780
 Db
 781 TGGTTCCAGGAGGATCAACAGCTTTGGCTGTGGAGGAGAGAACCGCTTGA 840
 QY
 781 TGGTTCCAGGAGGATCAACAGCTTTGGCTGTGGAGGAGAGAACCGCTTGA 840
 Db
 841 GTTTTCACTGCTGTGCTACCTATGAGGATGATACGGGAGCAGGTGATGGTTTCAAG 900
 QY
 841 GTTTTCACTGCTGTGCTACCTATGAGGATGATACGGGAGCAGGTGATGGTTTCAAG 900
 Db
 901 CTGGGCGCTGCTTTCCACCCAGCCCGCAGAGACCCAGTTCAGATTGTTTACATCAAG 960
 QY
 901 CTGGGCGCTGCTTTCCACCCAGCCCGCAGAGACCCAGTTCAGATTGTTTACATCAAG 960
 Db
 961 GCAATTCCTGGATTCTGCCAGAAATCTTTTGGAGCCCTTGTCCCATATATCAGTAGGATC 1020
 QY
 961 GCAATTCCTGGATTCTGCCAGAAATCTTTTGGAGCCCTTGTCCCATATATCAGTAGGATC 1020
 Db
 1021 TCAACTGGGACCAAAAGCCTTGTCTCCCTTGTCTCTCCACACTCTCTCTGGGCGCTC 1080
 QY
 1021 TCAACTGGGACCAAAAGCCTTGTCTCCCTTGTCTCTCTCCACACTCTCTCTGGGCGCTC 1080
 Db
 1081 TGGGGGTTCTGATGGGCGCTCC 1102
 QY
 1081 TGGGGGTTCTGATGGGCGCTCC 1102
 Db
 RESULT 3
 ABK31795
 ID ABK31795 standard; DNA; 2457 BP.
 XX AC
 XX ABK31795;
 XX AC
 XX 23-APR-2002 (first entry)
 XX
 XX DNA encoding novel human protease #52.
 XX
 XX Human; protease; cancer; immune-related disorder; cardiovascular disease;
 KW neuronal-associated disease; metabolic disorder; inflammatory disorder;
 KW nervous system disorder; sexual dysfunction; pain; mood disorder;
 KW hypertension; psychotic disorder; neurological disorder; dyskinesia;
 KW viral infection; human immunodeficiency virus; HIV; non-viral infection;
 KW ocular disease; cytostatic; gene; ds.
 XX
 XX Homo sapiens.
 XX OS
 XX WO200200860-A2.
 XX
 XX 03-JAN-2002.
 XX
 XX 26-JUN-2001; 2001WO-US020171.
 XX PF
 XX 26-JUN-2000; 2000US-0214047P.
 XX PR
 XX (SUGEN-) SUGEN INC.
 XX PA
 XX Plowman G, Whyte D, Sudarsanam S, Manning G, Caenepeel S;
 XX Charydczak G;
 XX
 XX WPI; 2002-139913/18.
 XX DR
 XX P-PSDB; AAU82753.
 XX
 XX Nucleic acids encoding novel human proteases, useful for use for

treating diseases and disorders such as cancers, immune-related diseases
 and disorders, cardiovascular disease (e.g. restenosis) and inflammatory
 disorders.
 XX
 PS Claim 26; Fig 1S8-1T; 313pp; English.
 XX
 The present invention relates to the isolation of novel human proteases,
 and the nucleic acids encoding them. The sequences of the invention are
 useful for treating diseases and disorders such as cancers (e.g. breast,
 colon, lung), immune-related diseases and disorders (e.g. inflammatory
 diseases and asthma), cardiovascular diseases (e.g. restenosis and
 coronary thrombosis), brain or neuronal-associated diseases, metabolic
 disorders (e.g. diabetes, obesity), inflammatory disorders (e.g.
 rheumatoid arthritis and psoriasis), central or peripheral nervous system
 diseases, migraines, pain, sexual dysfunction, mood disorders, attention
 disorders, cognitive disorders, hypotension, hypertension, psychotic
 disorders, neurological disorders (e.g. Alzheimer's disease, Parkinson's
 disease) and dyskinesias. The nucleic acids and polypeptides are also
 useful for treating viral infections caused by human immunodeficiency
 virus (HIV), and non-viral infections such as ocular disease (e.g.
 glaucoma) and macular degeneration. ABK31744-ABK31802 represent DNA
 sequences encoding for the novel human proteases of the invention
 XX
 QY Sequence 2457 BP; 377 A; 851 C; 797 G; 432 T; 0 U; 0 Other;
 SQ
 Query Match 65.0%; Score 716; DB 6; Length 2457;
 Best Local Similarity 99.8%; Pred. No. 0;
 Matches 816; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 128 ACTGGGGGGCGCTCAGCCCTCGGCGCCGCTCAATCGTGGGGGGCTCAAAACGCGAGCGGGCA 187
 QY
 110 ACTGGGGGGCGCTCAGCCCTCGGCGCCGCTCAATCGTGGGGGGCTCAAAACGCGAGCGGGCA 169
 Db
 188 CTGGGCTTGGCAAGTGAGCTGACCATGAGAGGTGGCCACATCTGCGGGGGTCTCCCTCA 247
 QY
 170 CTGGGCTTGGCAAGTGAGCTGACCATGAGAGGTGGCCACATCTGCGGGGGTCTCCCTCA 229
 Db
 248 TCGCCCGCTCTGGGGTCTCTCGGTGCTCACTGTTTTCATGAGATGGAGAGTGGAGC 307
 QY
 230 TCGCCCGCTCTGGGGTCTCTCGGTGCTCACTGTTTTCATGAGATGGAGAGTGGAGC 289
 Db
 308 CGCGCGCGAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 367
 QY
 290 CGCGCGCGAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 349
 Db
 368 CGCACACCGCGCAGTGGCGGCGCATCGTGGTGGCGGCGCAACTACAGCCAAAGTGGAGCTGG 427
 QY
 350 CGCACACCGCGCAGTGGCGGCGCATCGTGGTGGCGGCGCAACTACAGCCAAAGTGGAGCTGG 409
 Db
 428 GCGCGCGAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGC 487
 QY
 410 GCGCGCGAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGC 469
 Db
 488 CTGCTGCTGCGCGCGCGCTCACCGCTTCTGTGACCGGCAACCGGCTGCTGGGGCGCACCG 547
 QY
 470 CTGCTGCTGCGCGCGCGCTCACCGCTTCTGTGACCGGCAACCGGCTGCTGGGGCGCACCG 529
 Db
 548 GCTGGGGAGAGCTCCAGAGGAGAGTCTCTGCTCTCTCCCTGGGTGCTACAGAAAGTGG 607
 QY
 530 GCTGGGGAGAGCTCCAGAGGAGAGTCTCTGCTCTCTCCCTGGGTGCTACAGAAAGTGG 589
 Db
 608 AGCTAAGGCTGCTGGGGAGGCGCACCTGTCAATGTCTCTACAGCCAGCCCGGTCCCTTCA 667
 QY
 590 AGCTAAGGCTGCTGGGGAGGCGCACCTGTCAATGTCTCTACAGCCAGCCCGGTCCCTTCA 649
 Db
 668 ACCTCACTCTCCAGATATTGCCAGGAGTGTGTGTGCTGCTGCTACCCAGAGGGCGCGAGGG 727
 QY
 650 ACCTCACTCTCCAGATATTGCCAGGAGTGTGTGTGCTGCTGCTACCCAGAGGGCGCGAGGG 709
 Db
 728 ACACCTGCCAGGGTGAATCTGGGGGGCGCTCTGCTGTGTGTGTGTGTGTGTGTGTGTGTTC 787
 QY
 710 ACACCTGCCAGGGTGAATCTGGGGGGCGCTCTGCTGTGTGTGTGTGTGTGTGTGTGTGTTC 769
 Db

[illegible]

AA		Homo sapiens.	
OS			
XX			
XX	Key	Location/Qualifiers	
FH	CDS	1..768	
FT		/tag= a	
PT		/product= "secreted polypeptide"	
FT			
XX			
PN	WO200166690-A2.		
XX			
XX	13-SEP-2001.		
XX			
PF	05-MAR-2001; 2001WO-US007143.		
XX			
PR	06-MAR-2000; 2000US-0187107P.		
PR	13-MAR-2000; 2000US-0188916P.		
PR	03-OCT-2000; 2000US-0236874P.		
PR	03-OCT-2000; 2000US-0237846P.		
XX			
PA	(SMIK) SMITHKLINE BEECHAM CORP.		
PA	(SMIK) SMITHKLINE BEECHAM PLC.		
XX			
PI	Agarwal P, Murdoch PR, Rizvi SK, Smith RF, Xiang Z;		
XX			
DR	WFI; 2001-570768/64.		
DR	P-PSDE; AAG67514.		

[illegible]

PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity.

XX Claim 1; SEQ ID NO 6883; 103pp; English.

XX The invention relates to isolated polynucleotide (I) and polypeptide (II)
CC sequences. (I) is useful as hybridisation probes, polymerase chain
CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
CC and in recombinant production of (II). The polynucleotides are also used
CC in diagnostics as expressed sequence tags for identifying expressed
CC genes. (I) is useful in gene therapy techniques to restore normal
CC activity of (II) or to treat disease states involving (II). (II) is
CC useful for generating antibodies against it, detecting or quantitating a
CC polypeptide in tissue, as molecular weight markers and as a food
CC supplement. (II) and its binding partners are useful in medical imaging
CC of sites expressing (II). (I) and (II) are useful for treating disorders
CC involving aberrant protein expression or biological activities. The
CC polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. AAS64197-AAS94564 represent novel human diagnostic
CC coding sequences of the invention. Note: The sequence data for this
CC patent did not appear in the printed specification, but was obtained in
CC electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences

XX Sequence 2298 BP; 468 A; 774 C; 706 G; 350 T; 0 U; 0 Other;

Query Match 12.3%; Score 135; DB 5; Length 2298;
Best Local Similarity 100.0%; Pred. No. 8.6e-53;
Matches 135; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 811 TTTGGCTGTGACGAGAGAAACCGCCCTGGAGTTTTCATCTGTGGTACCTATGAGGCA 870
DB 445 TTTGGCTGTGACGAGAGAAACCGCCCTGGAGTTTTCATCTGTGGTACCTATGAGGCA 504
QY 871 TGGATACGGGAGCGAGTGTATGGTTTCAGAGCTGGGCTTCCACCCAGCCCCAG 930
DB 505 TGGATACGGGAGCGAGTGTATGGTTTCAGAGCTGGGCTTCCACCCAGCCCCAG 564
QY 931 AAGACCCAGTCAGAT 945
DB 565 AAGACCCAGTCAGAT 579

RESULT 13
ABT31936
ID ABT31936 standard; DNA; 1733 BP.

XX ABT31936;
XX
XX
XX
XX 01-MAY-2003 (first entry)
XX Human breast cancer / ovarian cancer related coding sequence #43.
DE Human; gene; ds; cytostatic; breast cancer; ovarian cancer.
KW Homo sapiens.
XX
XX WO2003000012-A2.
XX
XX 03-JAN-2003.
XX
XX 21-JUN-2002; 2002WO-US019773.
XX
XX 21-JUN-2001; 2001US-0300159P.
XX 27-JUN-2001; 2001US-0301351P.
XX
XX (MILL-) MILLENNIUM PHARM INC.

PI Veiby OP;

XX WPI; 2003-267848/26.
DR P-PSDB; ABJ37067.

XX Determining the presence of breast cancer in an individual, involves
PT using specific polynucleotide markers.

XX Disclosure; Page 192-193; 233pp; English.

XX The invention comprises a method for assessing whether a patient is
CC afflicted with breast cancer or ovarian cancer. The method involves the
CC use of specific DNA markers. The method of the invention is useful in the
CC detection and treatment of ovarian and breast cancer. DNA sequences
CC ABT31894 - ABT31949 encode human breast/ovarian cancer-related proteins
XX
XX Sequence 1733 BP; 311 A; 578 C; 500 G; 344 T; 0 U; 0 Other;

Query Match 6.1%; Score 67; DB 7; Length 1733;
Best Local Similarity 100.0%; Pred. No. 5e-21;
Matches 67; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GGACCCCTTCTCTGGGCGCATGGCCCAAGAGGGGGTCTCTGGGGCTGGGCAGCTGGGGGCT 60
DB 81 GGACCCCTTCTCTGGGCGCATGGCCCAAGAGGGGGTCTCTGGGGCTGGGCAGCTGGGGGCT 140
QY 61 GTGGCCA 67
DB 141 GTGGCCA 147

RESULT 14
AAF98698
ID AAF98698 standard; DNA; 1796 BP.

XX AAF98698;
XX
XX 02-JUL-2001 (first entry)
XX Human ovarian cancer cell expressed sequence 10798.

XX Human; ovarian cancer; identification; detection; characterisation;
KW tumour; kinase; marker; cytostatic; antisense gene therapy; ds.

XX Homo sapiens.

XX WO200118542-A2.

XX 15-MAR-2001.

XX 01-SEP-2000; 2000WO-US024199.

XX 03-SEP-1999; 99US-0152547P.

XX 16-MAR-2000; 2000US-0190347P.

XX 21-MAR-2000; 2000US-0191321P.

XX 31-MAY-2000; 2000US-0208382P.

XX 20-JUL-2000; 2000US-00220467.

XX (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.

XX Lee J, Thompsho P, Lillie J;

XX WPI; 2001-211428/21.

XX Detection, assessment, prevention and therapy of ovarian cancer,
PT comprises detecting changes in the expression of a variety of markers.

XX Claim 1; Page 1001-1002; 1198pp; English.

XX The present invention describes a method for assessing whether a patient
CC is afflicted with ovarian cancer by comparing: (1) the expression of a
CC marker (I) (see AAF98594 to AAF98730), in a patient sample; and (2) the
CC normal level of expression of (I) in a control non-ovarian cancer sample,

CC where a significant difference between the level of expression in (a) and
CC (b) is an indication that the patient is afflicted with ovarian cancer.
CC (1) have cytostatic activities and can be used in antisense gene therapy.
CC The method, compositions and kits from the present invention can be used
CC for: (1) assessing and treating ovarian cancer; (2) making isolated
CC hybridoma, which produces an antibody useful for ovarian cancer
CC assessment; and (3) inhibiting ovarian cancer in a patient. AAF98573 to
CC AAF98593 represent human kinase marker primers and probes which are used
CC in the exemplification of the present invention

XX Sequence 1796 BP; 327 A; 599 C; 511 G; 359 T; 0 U; 0 Other;
SQ

Query Match 6.1%; Score 67; DB 5; Length 1796;
Best Local Similarity 100.0%; Pred. No. 5e-21;
Matches 67; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GGGCCCTTGCTGGGCGCATGGCCAGAGGGGGTCTCTGGGGCTGGGCGAGCTGGGGGCT 60
DB 134 GGGCCCTTGCTGGGCGCATGGCCAGAGGGGGTCTCTGGGGCTGGGCGAGCTGGGGGCT 193
QY 61 GTGGCCA 67
DB 194 GTGGCCA 200

RESULT 15
AAF98720
ID AAF98720 standard; DNA; 1835 BP.
XX
AC AAF98720;
XX
DT 02-JUL-2001 (first entry)
XX
DE Human late stage ovarian tumour polynucleotide marker 28.
XX
KW Human; ovarian cancer; identification; detection; characterisation;
KW tumour; kinase; marker; cytostatic; antisense gene therapy; ds.
XX
OS Homo sapiens.
XX
FN WO200118542-A2.
XX
PD 15-MAR-2001.
XX
PF 01-SEP-2000; 2000WO-US024199.
XX
PR 03-SEP-1999; 99US-0152547P.
PR 16-MAR-2000; 2000US-0190347P.
PR 21-MAR-2000; 2000US-0191321P.
PR 31-MAY-2000; 2000US-0208382P.
PR 20-JUL-2000; 2000US-00220467.
XX
PA (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.

XX Lee J, Thompsho P, Lillie J;
XX WPI; 2001-211428/21.
DR
XX
PT Detection, assessment, prevention and therapy of ovarian cancer,
PT comprises detecting changes in the expression of a variety of markers.
XX
PS Claim 1; Page 1186-1187; 1198pp; English.

XX The present invention describes a method for assessing whether a patient
CC is afflicted with ovarian cancer by comparing: (1) the expression of a
CC marker (1) (see AAF98594 to AAF98730), in a patient sample; and (2) the
CC normal level of expression of (1) in a control non-ovarian cancer sample,
CC where a significant difference between the level of expression in (a) and
CC (b) is an indication that the patient is afflicted with ovarian cancer.
CC (1) have cytostatic activities and can be used in antisense gene therapy.
CC The method, compositions and kits from the present invention can be used
CC for: (1) assessing and treating ovarian cancer; (2) making isolated
CC hybridoma, which produces an antibody useful for ovarian cancer

CC assessment; and (3) inhibiting ovarian cancer in a patient. AAF98573 to
CC AAF98593 represent human kinase marker primers and probes which are used
CC in the exemplification of the present invention

XX Sequence 1835 BP; 309 A; 621 C; 527 G; 378 T; 0 U; 0 Other;
SQ

Query Match 6.1%; Score 67; DB 5; Length 1835;
Best Local Similarity 100.0%; Pred. No. 5e-21;
Matches 67; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GGGCCCTTGCTGGGCGCATGGCCAGAGGGGGTCTCTGGGGCTGGGCGAGCTGGGGGCT 60
DB 212 GGGCCCTTGCTGGGCGCATGGCCAGAGGGGGTCTCTGGGGCTGGGCGAGCTGGGGGCT 271
QY 61 GTGGCCA 67
DB 272 GTGGCCA 278

Search completed: February 25, 2004, 12:36:33
Job time : 478 secs

Thu Feb 26 13:50:22 2004

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 25, 2004, 12:18:11 ; Search time 2783 Seconds
(without alignments)
11824.694 Million cell updates/sec

Title: US-10-037-417-45

Perfect score: 1102
Sequence: 1 gggccctgtctctggcccat.....ggggtctgatggggcctcc 1102

Scoring table: OLIGO NUC
Gapop '60.0 , Gapext 60.0

Searched: 27513289 seqs, 14931090276 residues

Word size : 0

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

EST:*

1: em_estba.*

2: em_esthum.*

3: em_estin.*

4: em_estmu.*

5: em_estov.*

6: em_estpl.*

7: em_estro.*

8: em_htc.*

9: gb_est1.*

10: gb_est2.*

11: gb_est3.*

12: gb_est4.*

13: gb_est5.*

14: gb_estfun.*

15: em_estfun.*

16: em_estom.*

17: em_gss_hum.*

18: em_gss_inv.*

19: em_gss_pln.*

20: em_gss_vrt.*

21: em_gss_fun.*

22: em_gss_mam.*

23: em_gss_mus.*

24: em_gss_pro.*

25: em_gss_rod.*

26: em_gss_pig.*

27: em_gss_vri.*

28: gb_gss1.*

29: gb_gss2.*

SUMMARIES

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

Result No.	Score	Query Match	Length	DB ID	Description
C 1	466	42.3	537	10	AW450407
C 2	384	34.8	456	9	AI190509
C 3	135	12.3	930	13	BX436299
4	67	6.1	294	9	AA300017

67	6.1	318	12	BM828821	BM828821 K-EST0101
67	6.1	360	9	AI393077	AI393077 tg25f08.x
67	6.1	366	9	AI623099	AI623099 tu49f05.x
67	6.1	377	9	AI193435	AI193435 qe58e10.x
67	6.1	406	14	CB116948	CB116948 K-EST0161
67	6.1	409	9	AU298208	AU298208 AU298208
67	6.1	462	9	AI343968	AI343968 tg01405.x
67	6.1	472	9	AI761519	AI761519 w161402.x
67	6.1	479	9	AU298317	AU298317 AU298317
67	6.1	479	14	CB996389	CB996389 AGENCOURT
67	6.1	492	10	BF063417	BF063417 7n89e10.x
67	6.1	516	12	BM311864	BM311864 1963a10.Y
67	6.1	527	9	AI658798	AI658798 tu08g09.x
67	6.1	527	12	BG482000	BG482000 602527844
67	6.1	534	9	AI393068	AI393068 tg25e09.x
67	6.1	537	12	BM837284	BM837284 K-EST0113
67	6.1	537	12	BM837397	BM837397 K-EST0113
67	6.1	542	10	BE280845	BE280845 601155410
67	6.1	555	12	BM837589	BM837589 K-EST0113
67	6.1	570	12	BG479276	BG479276 602526495
67	6.1	570	14	CB125988	CB125988 K-EST0174
67	6.1	575	9	AI207222	AI207222 qf81h10.x
67	6.1	583	12	BG386531	BG386531 602455771
67	6.1	590	13	BQ292219	BQ292219 PM2-AN008
67	6.1	591	9	AU134797	AU134797 AU134797
67	6.1	594	10	BE281529	BE281529 601155121
67	6.1	604	9	AU298212	AU298212 AU298212
67	6.1	608	14	CD723955	CD723955 OJ28e07.Y
67	6.1	619	12	BM794592	BM794592 K-EST0075
67	6.1	625	9	AU134318	AU134318 AU134318
67	6.1	632	9	AU133801	AU133801 AU133801
67	6.1	648	9	AU138370	AU138370 AU138370
67	6.1	657	9	AI935459	AI935459 w084h02.x
67	6.1	659	9	AU139645	AI935459 AU139645
67	6.1	666	9	AU136193	AU136193 AU136193
67	6.1	670	9	AU121662	AU121662 AU121662
67	6.1	681	9	AU139870	AU139870 AU139870
67	6.1	683	13	BX505123	BX505123 DKFZp686J
67	6.1	688	9	AU136859	AU136859 AU136859
67	6.1	690	12	BG481221	BG481221 602528547

ALIGNMENTS

RESULT 1

AW450407/c

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

AW450407 537 bp mRNA linear EST 17-FEB-2000
UI-H-BI3-akn-g-11-0-UI-s1 NCI CGAP_Sub5 Homo sapiens cDNA clone
IMAGE:2735037 3', mRNA sequence.

AW450407 GI:6991183

EST.

Homo sapiens (human)

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 537)

NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.

National Cancer Institute, Cancer Genome Anatomy Project (CGAP),

Tumor Gene Index

Unpublished (1997)

Contact: Robert Strausberg, Ph.D.

Email: cgaps@mail.nih.gov

Oligo-dt track not found, Not I site shown in beginning of sequence

is likely internal to the message. cDNA library preparation: M.B.

Soares Lab Clone distribution: NCI-CGAP clone distribution

Information can be found through the I.M.A.G.E. Consortium/LML at:

www.bio.llnl.gov/bbrp/image/image.html

Seq primer: M13 Forward

POLYA=No.

Location/Qualifiers

```
source
1. .537
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:2735037"
/lab_host="DH10B (Life Technologies)"
/clone_lib="NCI_CGAP_Sub5"
/notes="Vector: pT73D-Pac (Pharmacia) with a modified
polylinker; Site 1: Not 1; Site 2: Eco RI; NCI CGAP Sub5
is a subtracted library derived from NCI CGAP Sub4. The
NCI CGAP Sub5 library had 3 million recombinants. A
single-stranded DNA preparation of NCI CGAP Sub4 was used
as a tracer in a subtractive hybridization with a driver
comprising the IMAGE pool (NCI CGAP Kid3 pool 1 LLM
3334-3337, 3682-3683, 3798-3803 (IMAGE Clonoids
1322376-1323911, 1456008-1456775, 1500552-1502855);
NCI CGAP Kid5 pool 1 LLM 3338-3342, 3722-3725, 3776-3778
(IMAGE Clonoids 1373912-1325831, 1471368-1472903,
1492104-1492255); NCI CGAP_Lu5 pool 1 LLM 3575-3582,
3851-3854 (IMAGE Clonoids
1414920-1417991, 1520904-1522439); NCI CGAP_G4 pool 1 LLM
3164-3167, 3716-3720, 3733-3735 (IMAGE Clonoids
1257096-1258631, 1469064-1470983, 1475592-1476743);
NCI CGAP_Pr22 pool 1 LLM 2457-2459, 2758-2759, 3062-3068
(IMAGE Clonoids 985608-986759, 1101192-1101959,
1217928-1220615); NCI CGAP_Col0 pool 1 LLM 2644-2653,
2871-2872 (IMAGE Clonoids
1057416-1061255, 1144584-1145351). (10% of the driver
population), plus a pool of 3,840 arrayed clones from
NCI CGAP Sub1 (IMAGE Clonoids 2708616-2710535) and
NCI_CGAP_Sub2 (IMAGE Clonoids 2710536-2712455) (10% of
the driver population), plus a pool of 11,136 clones from
NCI CGAP Sub3 (IMAGE Clonoids 2712456-2723591) (10% of the
driver population), plus a pool of 5,472 clones from
NCI CGAP Sub4 (IMAGE Clonoids 2723592-2728969) (70% of the
driver population). Subtraction was performed as
previously described (Bonald, Lennon & Soares (1996):
Normalization and Subtraction: Two Approaches To
Facilitate Gene Discovery. Genome Research 6, 791-806.
TAG TISSUE=colon
TAG_LIB=NCI CGAP_Col0
TAG_SEQ=AAACGG"

106 Db CTCTCCAGATATTGCCAGGATGCTGTGTGCTTACCCAGGGCGCGAGGACACT 47
734 QY GCCAGGGTACTCTGGGGGCCCCCTGCTGTGTGAGGAAGCGGCGC 779
46 Db GCCAGGGTACTCTGGGGGCCCCCTGCTGTGTGAGGAAGCGGCGC 1

RESULT 2
A1190509 456 bp mRNA linear EST 28-OCT-1998
Qd49f07.x1 Soares fetal heart NbH19W Homo sapiens cDNA clone
IMAGE:1732837 3, similar to SW:MCP6_MOUSE P21845 MAST CELL PROTEASE
6 PRECURSOR 1, mRNA sequence.
A1190509
A1190509.1 GI:3741718
EST. Homo sapiens (human)
Homo sapiens
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 456)
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgapsb@mail.nih.gov
This clone is available royalty-free through LML; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Insert Length: 712 Std Error: 0.00
Seq primer: -40UP from Gibco
High quality sequence stop: 452.
Location/Qualifiers
1. .456
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:1732837"
/sex="unknown"
/dev stage="19 weeks"
/lab host="DH10B (ampicillin resistant)"
/clone_lib="Soares fetal heart NbH19W"
/notes="Organ: heart; Vector: pT73D (Pharmacia) with a
modified polylinker; Site 1: Not 1; Site 2: Eco RI; 1st
strand cDNA was primed with a Not I - oligo(dT) primer [5',
TGTACCAATCTGAAGTGGAGCGCGGCACCTTTTCTTTTCTTTT 3'],
double-stranded cDNA was size selected, ligated to Eco RI
adapters (Pharmacia), digested with Not I and cloned into
the Not I and Eco RI sites of a modified pT73 vector
(Pharmacia). Library went through one round of
normalization to a Cot = 5. Library constructed by
M.Fatima Bonald. This library was constructed from the
same fetus as the fetal lung library, Soares fetal lung
NbH19W."

FEATURES
source
Query Match 34.8%; Score 384; DB 9; Length 456;
Best Local Similarity 100.0%; Pred. No. 1.6e-170; Mismatches 0; Indels 0; Gaps 0;
Matches 384; Conservative

379 QY GCAGTGGCGCCCATCGTGTGGCGCCCAACTACAGCCAAAGTGGAGCTGGGGCCGACCTG 438
384 Db GCAGTGGCGCCCATCGTGTGGCGCCCAACTACAGCCAAAGTGGAGCTGGGGCCGACCTG 325
439 QY GCCCTGCTGCGCTTGGCTTCACTCCCGCAGCTTGGGGCCCGCGCTGTGCTGTGCTG 498
324 Db GCCCTGCTGCGCTTGGCTTCACTCCCGCAGCTTGGGGCCCGCGCTGTGCTGTGCTG 265
499 QY CCCCAGGCTTACACCGCTTCTGTCACCGCAGCGCTGTGGCCACCGCTGTGGGAGAC 558
264 Db CCCCAGGCTTACACCGCTTCTGTCACCGCAGCGCTGTGGCCACCGCTGTGGGAGAC 205

ORIGIN
Query Match 34.8%; Score 384; DB 9; Length 456;
Best Local Similarity 100.0%; Pred. No. 1.6e-170; Mismatches 0; Indels 0; Gaps 0;
Matches 384; Conservative

379 QY GCAGTGGCGCCCATCGTGTGGCGCCCAACTACAGCCAAAGTGGAGCTGGGGCCGACCTG 438
384 Db GCAGTGGCGCCCATCGTGTGGCGCCCAACTACAGCCAAAGTGGAGCTGGGGCCGACCTG 325
439 QY GCCCTGCTGCGCTTGGCTTCACTCCCGCAGCTTGGGGCCCGCGCTGTGCTGTGCTG 498
324 Db GCCCTGCTGCGCTTGGCTTCACTCCCGCAGCTTGGGGCCCGCGCTGTGCTGTGCTG 265
499 QY CCCCAGGCTTACACCGCTTCTGTCACCGCAGCGCTGTGGCCACCGCTGTGGGAGAC 558
264 Db CCCCAGGCTTACACCGCTTCTGTCACCGCAGCGCTGTGGCCACCGCTGTGGGAGAC 205

ORIGIN
Query Match 42.3%; Score 466; DB 10; Length 537;
Best Local Similarity 100.0%; Pred. No. 2.5e-209; Mismatches 0; Indels 0; Gaps 0;
Matches 466; Conservative

314 QY CCGAGTGTGCTGCTGCTGGCGGTGCACTCCAGGACGGGCGCTCCGAGCGCGCACA 373
466 Db CCGAGTGTGCTGCTGCTGGCGGTGCACTCCAGGACGGGCGCTCCGAGCGCGCACA 407
374 QY CCGCGCAGTGGCGCGCATCGTGTGGCGCCCAACTACAGCCAAAGTGGAGCTGGGCGCG 433
406 Db CCGCGCAGTGGCGCGCATCGTGTGGCGCCCAACTACAGCCAAAGTGGAGCTGGGCGCG 347
434 QY ACCTGGCCCTGCTGGCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 493
346 Db ACCTGGCCCTGCTGGCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 287
494 QY GCCTGCCCCGCGCTTACACCGCTTCTGTCACCGCAGCGCTGCTGCTGCTGCTGCTG 553
286 Db GCCTGCCCCGCGCTTACACCGCTTCTGTCACCGCAGCGCTGCTGCTGCTGCTGCTG 227
554 QY GAGACGCTCCAGGAGGAGATTCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 613
226 Db GAGACGCTCCAGGAGGAGATTCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 167
614 QY GGCTGTGGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 673
166 Db GGCTGTGGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 107
674 QY CTCTCCAGATATTGCCAGGATGCTGTGTGCTGCTTACCCAGGGCGCGAGGACACT 733
```

db 226 AAGACCCAGTCAGAT 240

RESULT 4
AA300017
LOCUS
DEFINITION
EST12620 Uterus tumor I Homo sapiens cDNA 5' end similar to
Protasain, mRNA sequence.

ACCESSION
AA300017
VERSION
AA300017.1
KEYWORDS
GI:1952347

SOURCE
Homo sapiens (human)

ORGANISM
Homo sapiens

REFERENCE
AUTHORS
Bukayota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
1 (bases 1 to 294)
Adams,M.D., Kerlavage,A.R., Fleischmann,R.D., Fuldner,R.A.,
Bult,C.J., Lee,N.H., Kirkness,E.F., Weinstock,K.G., Gocayne,J.D.,
White,O., Sutton,G., Blake,J.A., Brandon,R.C., Man-Wai,C.,
Clayton,R.A., Cline,T.R., Cotton,M.D., Barle-Hughes,J., Fine,L.D.,
Pitgerald,L.M., Fitzhugh,W.M., Fritchman,J.L., Geoghegan,N.S.,
Glodde,K., Gnehm,C.L., Hanna,M.C., Hedblom,E., Hinkle,P.S.Jr.,
Kelley,A.M., Kelley,J.C., Liu,L.-I., Marmaros,S.M., Merrick,J.M.,
Moreno-Balancues,R.F., McDonald,L.A., Nguyen,D.T., Pelligrino,S.M.,
Phillips,C.A., Ryder,S.E., Scott,J.L., Saudek,D.M., Shirley,R.,
Small,K.V., Spriggs,T.A., Uterback,M.A., Coleman,I.A., Collins,E.J.,
Bednarek,D.P., Cao,L., Cepeda,M.A., Fischer,C., Hastings,G.A.,
Dimke,D., Feng,D.-F., Ferrie,A., Fischer,J., Hudson,P., Kim,A.K.,
He,W., Hu,J.S., Greene,J.M., Gruber,J., Li,H., Weissner,P.S., Olsen,H.,
Kozak,D.L., Kunsch,C., Hungjun,J., Li,H., Ruben,S.M.,
Raymond,L., Wei,Y.F., Wang,J., Xu,C., Yu,G.L., Ruben,S.M.,
Dillon,P.J., Fannon,M.R., Rosen,C.A., Haseltine,W.A., Fields,C.,
Fraser,C.M. and Venter,J.C.

TITLE
Initial assessment of human gene diversity and expression patterns
based upon 83 million nucleotides of cDNA sequence

JOURNAL
NATURE 377 (6547 Suppl), 3-174 (1995)

MEDLINE
96026280

PUBMED
7566098

COMMENT
Other ESTs: EST12621 THCI76739
Contact: Kerlavage, AR
Bioinformatics
The Institute for Genomic Research
9712 Medical Center Drive, Rockville, MD 20850 USA
Tel: 3018699056
Fax: 3018699423
Email: arkerlav@tigr.org
For clone availability, additional sequence and expression
information related to this EST, please check the TIGR Human Gene
Index (<http://www.tigr.org/tdb/hgi/hgi.html>)
Seq primer: M13 Reverse.
Location/Qualifiers
1..294
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="ATCC (inhost):108171"
/db_xref="taxon:9606"
/dev_stage="adult"
/clone_lib="Uterus tumor I"
/note="Organ: uterus; Vector: pBluescript SK-; Site_1:
EcoRI; Site_2: XhoI"

FEATURES
source
6.1%; Score 67; DB 9; Length 294;
Query Match
Best Local Similarity 100.0%; Pred. No. 1.9e-20;
Matches 67; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

ORIGIN
1 GGGCCCTTCTCTGGGCCCATGGCCAGAGGGGGTCTCTGGGGCTTGGCAGCTGGGGCT 60
158 GGGCCCTTCTCTGGGCCCATGGCCAGAGGGGGTCTCTGGGGCTTGGCAGCTGGGGCT 120
61 GTGGCCA 67

AI93077	AI93077	360 bp	mRNA	linear	EST 30-MAR-1999
LOCUS	tg25f08.x1 NCI CGAP CLL1 Homo sapiens cDNA clone IMAGE:2109831 3'				
DEFINITION	similar to SW:FS8_HUMAN Q16651 PROSTASIN PRECURSOR ; mRNA sequence.				
ACCESSION	AI93077	GI:4222624			
VERSION	AI93077				
KEYWORDS	EST.				
SOURCE	Homo sapiens (human)				
ORGANISM	Homo sapiens				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.				
AUTHORS	1 (bases 1 to 360) NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap. National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index Unpublished (1997)				
JOURNAL	Contact: Robert Strausberg, Ph.D. Email: cgapsb@mail.nih.gov				
COMMENT	Tissue Procurement: Ash Alizadeh, John Byrd, M.D., Mike Grever, M.D., Louis M. Staudt, M.D., Ph.D. cDNA Library Prepared by: M. Bento Soares, Ph.D. cDNA Library Arrayed by: Greg Lennon, Ph.D. DNA Sequencing by: Washington University Genome Sequencing Center Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: www-bio.llnl.gov/bbrp/image/image.html Insert Length: 691 Std Error: 0.00 Seq primer: -40UP from Gibco High quality sequence stop: 334. Location/Qualifiers 1..360 /organism="Homo sapiens" /mol_type="mRNA" /db_xref="taxon:9606" /clone="IMAGE:2109831" /tissue_type="B-cell, chronic lymphocytic leukemia" /lab_host="DH10B" /notes="Vector: pT7T3D-Fac (Pharmacia) with a modified polylinker; Site 1: Not I; Site 2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5' TGTTACCATTCTGAAGTGAGGCGGCATTCGTCTTTTTTTTTTTTTTTTTTTT T 3']; double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT7T3 vector. The library is normalized, and was constructed by Bento Soares and M.Patima Bonaldo."				
FEATURES					
source					
ORIGIN					
Query Match	6.1%; Score 67;	DB 9;	Length 360;		
Best Local Similarity	100.0%;	Fred.No. 1.9e-20;	Mismatches 0;	Gaps 0;	
Matches	67;	Conservative			
Qy	1 GGGCCCTGTCTCTGGGCCATGCCCAGAAAGGGGGTCTTGGGCCCTGGGACGTGGGGCT 60				
Db	127 GGGCCCTGTCTCTGGGCCATGGCCCAAGGGGGTCTTGGGCCCTGGGACGTGGGGCT 186				
Qy	61 GTGGCCA 67				
Db	187 GTGGCCA 193				
RESULT 7					
LOCUS	AI623099	366 bp	mRNA	linear	EST 15-DEC-1999
DEFINITION	tu49f05.x1 NCI CGAP Pr28 Homo sapiens cDNA clone IMAGE:2254401 3' similar to SW:FS8_HUMAN Q16651 PROSTASIN PRECURSOR ; mRNA sequence.				
ACCESSION	AI623099				
VERSION	AI623099.1	GI:4648024			
KEYWORDS	EST.				
SOURCE	Homo sapiens (human)				

218 GTGGCCA 224

RESULT 5
 BM828821
 LOCUS
 DEFINITION
 ACCESSION
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 318)
 Kim,N.S., Hahn,Y., Oh,J.H., Lee,J.Y., Ahn,H.Y., Chu,M.Y., Kim,M.R.,
 Oh,K.J., Cheong,J.E., Sohn,H.Y., Kim,J.M., Park,H.S., Kim,S. and
 Kim,Y.S.
 21C Frontier Korean EST Project 2001
 Unpublished (2002)
 Contact: Kim YS
 Genome Research Center
 Korea Research Institute of Bioscience & Biotechnology
 52 Eoeun-dong Yuseong-gu, Daejeon 305-333, South Korea
 Tel: +82-42-860-4470
 Fax: +82-42-860-4409
 Email: yongung@mail.kribb.re.kr
 Plate: 51, row: E column: 07
 High quality sequence stop: 318.
 Location/Qualifiers
 1. 318
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clones="S9SNU601-51-E07"
 /sex="M"
 /tissue_type="Ascites"
 /cell_type="Epithelial"
 /cell_line="SNU-601"
 /lab_host="Top10F"
 /clone_lib="S9SNU601"
 /note="Organ: Stomach; Vector: pME18-FL3; Site 1: XhoI;
 Site 2: XhoI; The poly (A)+ RNA was dephosphorylated with
 bacterial alkaline phosphatase (BAP) and then decapped
 with tabacco acid pyrophosphatase (TAP). The decapped
 intact mRNA was ligated with DNA-RNA linker including Sfil
 site by treatment of T4 RNA ligase and the first strand
 cDNA was synthesized with Superscript II using Sfil
 oligo-dT primer. After first strand synthesis, RNA was
 degraded by NaOH treatment and cDNA was amplified by PCR
 reaction. The PCR products were digested with Sfil and
 cloned into DraIII- digested pME18-FL3 vector. The
 obtained cDNA vectors were used for transformation of
 competent cells E. coli Top10F, by electroporation method.
 The cDNA libraries constructed by this method are
 full-length enriched cDNA library."

ORIGIN
 Query Match 6.1%; Score 67; DB 12; Length 318;
 Best Local Similarity 100.0%; Pred. No. 1.9e-20;
 Matches 67; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 0;
 Qy 1 GGGGCCCTGTCTGGGCCATGGCCAGAGGGGGTCTCTGGGGCTTGGGCGAGCTGGGGGCT 60
 |||||
 Db 207 GGGGCCCTGTCTGGGGCATGGCCAGAGGGGGTCTCTGGGGCTTGGGCGAGCTGGGGGCT 266
 |||||
 Qy 61 GTGGCCA 67
 |||||
 Db 267 GTGGCCA 273

RESULT 6

Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: c9pbs-remail.n.h.gov
This clone is available royalty-free through LNL ; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Insert Length: 763 Std Error: 0.00
Seq primer: -40UP from Gibco
High quality sequence stop: 356.
Location/Qualifiers
1. .377
/organism="Homo sapiens"
/mol_type="RNA"
/db_xref="taxon:9606"
/clone="IMAGE:1743210"
/dev stage="19 weeks"
/lab host="DH10B (ampicillin resistant)"
/clone lib="Soares fetal lung NDH19W"
/note="Organ: lung; Vector: pT7T3D (Pharmacia) with a
modified polylinker; Site 1: Not I; Site 2: Eco RI; 1st
strand cDNA was primed with a Not I - oligo(dT) primer
[5'-TGTTACCATCATGAGTGGAGCGGCCCAATTTTTTTTTTTT-3'],
double-stranded cDNA was size selected, ligated to Eco RI
adapters (Pharmacia), digested with Not I and cloned into
the Not I and Eco RI sites of a modified pT7T3 vector
(Pharmacia). Library went through one round of
normalization to a Cot = 5. Library constructed by Bento
Soares and M.Fatima Bonaldo. This library was constructed
from the same fetus as the fetal heart library, Soares
fetal heart NDH19W."

	ORIGIN
Query Match	6.1%; Score 67; DB 9; Length 377;
Best Local Similarity	100.0%; Pred.No. 1.9e-20;
Matches 67; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	
QY 1	GCGCCCTTGTCCTGGGCCATGCCCCAGACAGGGGGTCTCTGGGGCTGGGCACGTGGGGGCT 60
Ddb 153	GCGCCCTTGTCCTGGGCCATGCCCCAGACAGGGGGTCTCTGGGGCTGGGCACGTGGGGGCT 212
QY 61	GTGGCCA 67
Ddb 213	GTGGCCA 219

RESULT 9				
CB116948	406 bp	mRNA	linear	EST 28-JAN-2003
LOCUS	K-EST0161920 L7N800102 Homo sapiens cDNA clone L7N800102-4-C01 5',			
DEFINITION	mRNA sequence.			
ACCESSION	CB116948			
VERSION	CB116948.1			
KEYWORDS	GI:27942755			
SOURCE	EST.			
ORGANISM	Homo sapiens (human)			
	Homo sapiens			
	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;			
	Mammalia; Eutheria; Primates; Catarrhini; Homiinae; Homo.			
REFERENCE	1 (bases 1 to 406)			
AUTHORS	Kim,N.S., Hahn,Y., Oh,J.H., Lee,J.Y., Ahn,H.Y., Chu,M.Y., Kim,M.R., Oh,K.J., Cheong,J.E., Sohn,H.Y., Kim,J.M., Park,H.S., Kim,S. and Kim,Y.S.			
TITLE	21C Frontier Korea EST Project 2001			
JOURNAL	Unpublished (2002)			
COMMENT	Contact: Kim YS			
	Genome Research Center			
	Korea Research Institute of Bioscience & Biotechnology			
	52 Eoeun-dong Yuseong-gu, Daejeon 305-333, South Korea			
	Tel: +82-42-860-4470			
	Fax: +82-42-860-4409			
	Email: yongsung@mail.kribb.re.kr			
	Plate: 4 row: C column: 01			
	High quality sequence stop: 406.			
FEATURES	Location/Qualifiers			
	1..406			
	source			

ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 366)
NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Michael J. Brownstein, M.D., Ph.D., Michael R.
Emert-Buck, M.D., Ph.D.
cDNA Library Preparation: M. Bento Soares, Ph.D.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html
Insert Length: 630 Std Error: 0.00
Seq Primer: -40UP from Gibco
High quality sequence stop: 350
POLYA=No.

FEATURES
source
1..366
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:2254401"
/sex="male"
/dev_stage="adult"
/lab_host="DH10B"
/clone_lib="NCI_CGAP_Pr28"
/note="Organ: prostate; Vector: pT73D-Pac (Pharmacia)
with a modified polylinker; Plasmid DNA from the
normalized library NCI CGAP Pr22 was prepared, and ss
circles were made in vitro. Following HAP purification,
this DNA was used as tracer in a subtractive hybridization
reaction. The driver was PCR-amplified cDNAs from a pool
of 5,000 clones made from the same library (cloneIDs
985608-986759, 110192-1101959, and 1217928-1220615).
Subtracted by Bento Soares and M. Fatima Bonaldo."

```

ORIGIN
Query Match      6.1%; Score 67; DB 9; Length 366;
Best Local Similarity 100.0%; Pred. No. 1.9e-20;
Matches 67; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1  GGGCCCTTGTCCTGGGCCATGCCCCAGAGGGGGTCTCTGGGGCTCTGGCAGCTGGGGGGCT 60
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db      153 GGGCCCTTGTCCTGGGCCATGCCCCAGAGGGGGTCTCTGGGGCTCTGGCAGCTGGGGGGCT 212

QY      61  GTGGCCA 67
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db      213 GTGGCCA 219

RESULT 8
AII193435
LOCUS      qes8e10.x1 Soares fetal lung NBHL19W Homo sapiens cDNA clone
DEFINITION IMAGE:1743210 3', similar to SW:PS8_HUMAN Q16651 PROSTASIN
PRECUSOR ?; mRNA sequence.
ACCESSION AII193435
VERSION    AII193435.1 GI:3744644
KEYWORDS   EST.
SOURCE     Homo sapiens (human)
ORGANISM   Homo sapiens
REFERENCE  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS    Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
TITLE      1 (bases 1 to 377)
REFERENCE  NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
AUTHORS    National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
TITLE      Tumor Gene Index

```

ORIGIN	Query Match	Best Local Similarity	Score	DB	Length	Indels	Gaps
<pre> /organism="Homo sapiens" /mol_type="mRNA" /db_xref="taxon:9606" /clone="L7N800102-4-C01" /sex="M" /lab_host="Top10F" /notes="Organ: Liver; Vector: pCNS-D2; Site 1: EcoRI; Site 2: NotI; The poly (A)+ RNA was dephosphorylated with bacterial alkaline phosphatase (BAP) and then decapped with tobacco acid pyrophosphatase (TAP). The decapped intact mRNA was ligated with DNA-RNA linker including EcoRI site by treatment of T4 RNA ligase and the first strand cDNA was synthesized from oligo dt-selected mRNA by priming with dt-tailed vector. The dt-tailed vector was adjusted to have about 60nt. The cDNA vector was circularized with E. coli DNA ligase after digestion of EcoRI which site is also included in vector. An RNA strand converted to a DNA strand by Okayama-Berg method. The obtained cDNA vectors were used for transposition of competent cells E. coli Top10F by electroporation method. The cDNA libraries constructed by this method are full-length enriched cDNA library." </pre>	6.1%;	100.0%;	0;	Indels	0;	Gaps	0;
<pre> 1 GGGCCCTTGTCTCTGGCCATGCCCAGAGGGGGTCTGGGGCTGGGCAGCTGGGGCT 60 Db 213 GGGCCCTTGTCTCTGGCCATGCCCAGAGGGGGTCTGGGGCTGGGCAGCTGGGGCT 272 QY 61 GTGGCCA 67 Db 273 GTGGCCA 279 </pre>	6.1%;	100.0%;	0;	Indels	0;	Gaps	0;
<pre> RESULT 11 AI343968 LOCUS DEFINITION 462 bp mRNA linear EST 08-APR-1999 tcol0d05.x1 NCI CGAP Col6 Homo sapiens cDNA clone IMAGE:2062569 3' similar to TR:Q16651 PROSTASIN.; mRNA sequence. ACCESSION AI343968.1 GI:4081174 VERSION AI343968 KEYWORDS Homo sapiens (human) SOURCE Homo sapiens ORGANISM Homo sapiens Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo. 1 (bases 1 to 462) REFERENCE 1 NCI-CCAP http://www.ncbi.nlm.nih.gov/ncicgap. AUTHORS National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index JOURNAL Unpublished (1997) COMMENT Contact: Robert Strausberg, Ph.D. Email: cgaps@emil.nih.gov Tissue Procurement: Ilan Kirsch, M.D., Michael R. Emmert-Buck, M.D., Ph.D. cDNA Library Preparation: M. Bento Soares, Ph.D. cDNA Library Arrayed by: Greg Lennon, Ph.D. DNA Sequencing by: Washington University Genome Sequencing Center Clone distribution: NCI-CCAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: www.biol.llnl.gov/bbrp/image/image.html Insert Length: 586 Std Error: 0.00 Seq primer: -40UP from Gibco High quality sequence stop: 456. Location/Qualifiers 1. 462 /organism="Homo sapiens" /mol_type="mRNA" /db_xref="taxon:9606" /clone="IMAGE:2062569" /tissue_type="colon tumor, RER+" /lab_host="DH10B" /clone_lib="NCI CGAP Col6" /notes="Organ: colon; Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker; Site 1: Not I; Site 2: Eco RI; Plasmid DNA from the normalized library NCI CGAP Col6 was prepared, and ss circles were made in vitro. Following HAP purification, this DNA was used as tracer in a subtractive hybridization reaction. The driver was PCR-amplified cDNAs from a pool of 5,000 clones made from the same library (cloneids 1057416-1061255, and 1144584-1145351). Subtraction by Bento Soares and M. Fatima Bonaldo. " </pre>	6.1%;	100.0%;	0;	Indels	0;	Gaps	0;
<pre> Query Match Best Local Similarity 100.0%; Pred. No. 2e-20; Matches 67; Conservative 0; Mismatches 0; Indels 0; Gaps 0; </pre>	6.1%;	100.0%;	0;	Indels	0;	Gaps	0;
<pre> QY 1 GGGCCCTTGTCTCTGGCCATGCCCAGAGGGGGTCTGGGGCTGGGCAGCTGGGGCT 60 Db 213 GGGCCCTTGTCTCTGGCCATGCCCAGAGGGGGTCTGGGGCTGGGCAGCTGGGGCT 272 QY 61 GTGGCCA 67 Db 273 GTGGCCA 279 </pre>	6.1%;	100.0%;	0;	Indels	0;	Gaps	0;
<pre> RESULT 10 AI343968 LOCUS DEFINITION 409 bp mRNA linear EST 08-MAY-2003 AU298208 male adult skin, full-length enriched chimpanzee cDNA Library Pan troglodytes versus cDNA clone PstA2814 5' similar to human RefSeq mRNA NM_002773, mRNA sequence. ACCESSION AU298208 VERSION AU298208.1 GI:29532507 KEYWORDS EST. SOURCE Pan troglodytes versus ORGANISM Pan troglodytes Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Euthera; Primates; Catarrhini; Hominidae; Pan. 1 (bases 1 to 409) REFERENCE 1 Sakate,R., Osada,N., Hida,M., Sugano,S., Hayasaka,I., Shimohira,N., Yanagi,S., Suto,Y., Hashimoto,K. and Hirai,M. Analysis of 5'-end sequences of chimpanzee cDNAs Genome Res. 13 (5), 1022-1026 (2003) JOURNAL 22612989 MEDLINE 12727913 PUBMED Contact: Momoki Hirai Department of Integrated Biosciences The University of Tokyo, Graduate School of Frontier Sciences 5-1-5 Kashiwanoha, Kashiwa, Chiba 277-8562, Japan Tel: 81-4-7136-3688 Fax: 81-4-7136-3687 Email: mhiraie@k.u-tokyo.ac.jp. Location/Qualifiers 1. 409 /organism="Pan troglodytes versus" /mol_type="mRNA" /sub_species="verus" /db_xref="taxon:37012" /clone="PstA2814" /sex="male" </pre>	6.1%;	100.0%;	0;	Indels	0;	Gaps	0;
<pre> Query Match Best Local Similarity 100.0%; Pred. No. 1.9e-20; Matches 67; Conservative 0; Mismatches 0; Indels 0; Gaps 0; </pre>	6.1%;	100.0%;	0;	Indels	0;	Gaps	0;
<pre> QY 1 GGGCCCTTGTCTCTGGCCATGCCCAGAGGGGGTCTGGGGCTGGGCAGCTGGGGCT 60 Db 209 GGGCCCTTGTCTCTGGCCATGCCCAGAGGGGGTCTGGGGCTGGGCAGCTGGGGCT 268 QY 61 GTGGCCA 67 Db 269 GTGGCCA 275 </pre>	6.1%;	100.0%;	0;	Indels	0;	Gaps	0;
<pre> RESULT 10 AI298208 LOCUS DEFINITION 409 bp mRNA linear EST 08-MAY-2003 AU298208 male adult skin, full-length enriched chimpanzee cDNA Library Pan troglodytes versus cDNA clone PstA2814 5' similar to human RefSeq mRNA NM_002773, mRNA sequence. ACCESSION AU298208 VERSION AU298208.1 GI:29532507 KEYWORDS EST. SOURCE Pan troglodytes versus ORGANISM Pan troglodytes Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Euthera; Primates; Catarrhini; Hominidae; Pan. 1 (bases 1 to 409) REFERENCE 1 Sakate,R., Osada,N., Hida,M., Sugano,S., Hayasaka,I., Shimohira,N., Yanagi,S., Suto,Y., Hashimoto,K. and Hirai,M. Analysis of 5'-end sequences of chimpanzee cDNAs Genome Res. 13 (5), 1022-1026 (2003) JOURNAL 22612989 MEDLINE 12727913 PUBMED Contact: Momoki Hirai Department of Integrated Biosciences The University of Tokyo, Graduate School of Frontier Sciences 5-1-5 Kashiwanoha, Kashiwa, Chiba 277-8562, Japan Tel: 81-4-7136-3688 Fax: 81-4-7136-3687 Email: mhiraie@k.u-tokyo.ac.jp. Location/Qualifiers 1. 409 /organism="Pan troglodytes versus" /mol_type="mRNA" /sub_species="verus" /db_xref="taxon:37012" /clone="PstA2814" /sex="male" </pre>	6.1%;	100.0%;	0;	Indels	0;	Gaps	0;
<pre> Query Match Best Local Similarity 100.0%; Pred. No. 1.9e-20; Matches 67; Conservative 0; Mismatches 0; Indels 0; Gaps 0; </pre>	6.1%;	100.0%;	0;	Indels	0;	Gaps	0;
<pre> QY 1 GGGCCCTTGTCTCTGGCCATGCCCAGAGGGGGTCTGGGGCTGGGCAGCTGGGGCT 60 Db 209 GGGCCCTTGTCTCTGGCCATGCCCAGAGGGGGTCTGGGGCTGGGCAGCTGGGGCT 268 QY 61 GTGGCCA 67 Db 269 GTGGCCA 275 </pre>	6.1%;	100.0%;	0;	Indels	0;	Gaps	0;
<pre> RESULT 10 AI298208 LOCUS DEFINITION 409 bp mRNA linear EST 08-MAY-2003 AU298208 male adult skin, full-length enriched chimpanzee cDNA Library Pan troglodytes versus cDNA clone PstA2814 5' similar to human RefSeq mRNA NM_002773, mRNA sequence. ACCESSION AU298208 VERSION AU298208.1 GI:29532507 KEYWORDS EST. SOURCE Pan troglodytes versus ORGANISM Pan troglodytes Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Euthera; Primates; Catarrhini; Hominidae; Pan. 1 (bases 1 to 409) REFERENCE 1 Sakate,R., Osada,N., Hida,M., Sugano,S., Hayasaka,I., Shimohira,N., Yanagi,S., Suto,Y., Hashimoto,K. and Hirai,M. Analysis of 5'-end sequences of chimpanzee cDNAs Genome Res. 13 (5), 1022-1026 (2003) JOURNAL 22612989 MEDLINE 12727913 PUBMED Contact: Momoki Hirai Department of Integrated Biosciences The University of Tokyo, Graduate School of Frontier Sciences 5-1-5 Kashiwanoha, Kashiwa, Chiba 277-8562, Japan Tel: 81-4-7136-3688 Fax: 81-4-7136-3687 Email: mhiraie@k.u-tokyo.ac.jp. Location/Qualifiers 1. 409 /organism="Pan troglodytes versus" /mol_type="mRNA" /sub_species="verus" /db_xref="taxon:37012" /clone="PstA2814" /sex="male" </pre>	6.1%;	100.0%;	0;	Indels	0;	Gaps	

```

QY 1 GGGCCCTTGTCTGGCCATGGCCAGAGGGGTCTCTGGGGCTGGGCAGCTGGGGCT 60
Db 127 GGGCCCTTGTCTGGCCATGGCCAGAGGGGTCTCTGGGGCTGGGCAGCTGGGGCT 186
QY 61 GTGGCCA 67
Db 187 GTGGCCA 193

RESULT 12
AI761519 472 bp mRNA linear EST 21-DEC-1999
LOCUS similar02.xl NCI CGAP Col6 Homo sapiens cDNA clone IMAGE:2394723 3'
DEFINITION similar to SW:FS88_HUMAN Q16651 PROSTASIN PRECURSOR ;, mRNA
AUTHORS sequence.
ACCESSION AI761519.1 GI:5177186
VERSION AI761519
KEYWORDS Homo sapiens (human)
SOURCE EST.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 472)
AUTHORS Sakate,R., Osada,N., Hida,M., Sugano,S., Hayasaka,I., Shimohira,N.,
Yanagi,S., Suto,Y., Hashimoto,K. and Hirai,M.
TITLE Analysis of 5'-end sequences of chimpanzee cDNAs
JOURNAL Genome Res. 13 (5), 1022-1026 (2003)
COMMENT 22612989
PUBMED 12727913
CONTACT: Momoki Hirai
Department of Integrated Biosciences
The University of Tokyo, Graduate School of Frontier Sciences
5-1-5 Kashiwanoha, Kashiwa, Chiba 277-8562, Japan
Tel: 81-4-7136-3688
Fax: 81-4-7136-3687
Email: mhiraik@k.u-tokyo.ac.jp.

FEATURES
source
1..472
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:2394723"
/tissue_type="colon tumor, RER+"
/lab_host="DH10B"
/clone_lib="NCI_CGAP_Col6"
/note="Organ: colon; Vector: pVT73D-Pac (Pharmacia) with a
modified polylinker; Site 1: Not 1; Site 2: Eco RI;
Plasmid DNA from the normalized library NCI_CGAP_Col6 was
prepared, and as circles were made in vitro. Following HAP
purification, this DNA was used as tracer in a subtractive
hybridization reaction. The driver was PCR-amplified cDNAs
from a pool of 5,000 clones made from the same library
(cloneIDs 1057416-1061255, and 1144584-1145351).
Subtraction by Bento Soares and M. Fatima Ronaldo. "

ORIGIN
Query Match 6.1%; Score 67; DB 9; Length 472;
Best Local Similarity 100.0%; Pred. No. 2e-20;
Matches 67; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GGGCCCTTGTCTGGCCATGGCCAGAGGGGTCTCTGGGGCTGGGCAGCTGGGGCT 60
Db 102 GGGCCCTTGTCTGGCCATGGCCAGAGGGGTCTCTGGGGCTGGGCAGCTGGGGCT 161
QY 61 GTGGCCA 67
Db 162 GTGGCCA 168

RESULT 13
AI761519 479 bp mRNA linear EST 08-MAY-2003
LOCUS similar02.xl NCI CGAP Col6 Homo sapiens cDNA clone PstA5634 5' similar to
DEFINITION human RefSeq mRNA NM_002773, mRNA sequence.
ACCESSION AU298317
VERSION AU298317.1 GI:29532616
KEYWORDS EST.
SOURCE Pan troglodytes verus
ORGANISM Pan troglodytes verus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.
REFERENCE 1 (bases 1 to 479)
AUTHORS Sakate,R., Osada,N., Hida,M., Sugano,S., Hayasaka,I., Shimohira,N.,
Yanagi,S., Suto,Y., Hashimoto,K. and Hirai,M.
TITLE Analysis of 5'-end sequences of chimpanzee cDNAs
JOURNAL Genome Res. 13 (5), 1022-1026 (2003)
COMMENT 22612989
PUBMED 12727913
CONTACT: Momoki Hirai
Department of Integrated Biosciences
The University of Tokyo, Graduate School of Frontier Sciences
5-1-5 Kashiwanoha, Kashiwa, Chiba 277-8562, Japan
Tel: 81-4-7136-3688
Fax: 81-4-7136-3687
Email: mhiraik@k.u-tokyo.ac.jp.

FEATURES
source
1..479
/organism="Pan troglodytes verus"
/mol_type="mRNA"
/db_xref="taxon:37012"
/clone="PstA5634"
/sex="male"
/tissue_type="skin"
/dev_stage="adult"
/clone_lib="male adult skin, full-length enriched
chimpanzee cDNA library"

ORIGIN
Query Match 6.1%; Score 67; DB 9; Length 479;
Best Local Similarity 100.0%; Pred. No. 2e-20;
Matches 67; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GGGCCCTTGTCTCTGGCCATGGCCAGAGGGGTCTCTGGGGCTGGGCAGCTGGGGCT 60
Db 212 GGGCCCTTGTCTCTGGCCATGGCCAGAGGGGTCTCTGGGGCTGGGCAGCTGGGGCT 271
QY 61 GTGGCCA 67
Db 272 GTGGCCA 278

RESULT 14
CB996389 479 bp mRNA linear EST 01-MAY-2003
LOCUS AGENCOURT_13621004 NIH_MGC_148 Homo sapiens cDNA clone
DEFINITION IMAGE:30331484 5', mRNA sequence.
ACCESSION CB996389
VERSION CB996389.1 GI:30290909
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 479)
AUTHORS NIH-MGC http://mgi.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabs-r@mail.nih.gov
Tissue Procurement: Dr. Stefan Hansson

```

cDNA Library Preparation: Michael J. Brownstein (NHGRI) with help and advice from Piero Carninci (RIKEN)
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
Plate: NDAW346 row: P column: 21
High quality sequence stop: 478.
Location/Qualifiers
1. .479
source

FEATURES

/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:30331484"
/tissue_type="pre-eclamptic placenta"
/lab_host="DH10B TonA"
/clone_lib="NIH MGC 148"
/note="Organ: placenta; Vector: pBluescriptR; Site 1: all-XhoI; Site 2: BamH; Library is oligo-dT primed and directionally cloned using primer 5'-TTTTTTTTTTTIVN-3', size-selected for average insert size 2.3 kb and normalized to ROT 5. This is a primary library enriched for full-length clones and constructed using the Cap-trapper method (Carninci, in preparation). Library constructed by M. Brownstein (NIH/NHGRI, National Institutes of Health). Note: this is a NIH_MGC Library."

ORIGIN

Query Match 6.1%; Score 67; DB 14; Length 479;
Best Local Similarity 100.0%; Pred. No. 2e-20;
Matches 67; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GGCCCTTGTCTCGGCCCATGGCCAGAGGGGGTCTCTGGGGCTGGGAGCTGGGGGCT 60
Db 239 GGCCCTTGTCTCGGCCCATGGCCAGAGGGGGTCTCTGGGGCTGGGAGCTGGGGGCT 298
QY 61 GTGGCCA 67
Db 299 GTGGCCA 305

RESULT 15

BF063417
LOCUS 7h89e10.x1 NCI CGAP Col6 Homo sapiens cDNA clone IMAGE:3323178 3'
DEFINITION similar to SW:FS8_HUMAN Q16651 PROSTASIN PRECURSOR ;, mRNA
sequence.
ACCESSION BF063417 GI:10822327
VERSION BF063417.1
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE 1 (bases 1 to 492)
AUTHORS NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgaps@mail.nih.gov
Tissue Procurement: Ilan Kirsch, M.D., Michael R. Emmert-Buck, M.D., Ph.D.
cDNA Library Preparation: M. Bento Soares, Ph.D.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL, send email to: info@image.llnl.gov
Seq primer: -40UP from Gibco.
Location/Qualifiers

FEATURES

1. .492
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:3323178"
/tissue_type="colon tumor, RER+"
/lab_host="DH10B"
/clone_lib="NCI CGAP Col6"
/note="Organ: colon; Vector: pT73D-Pac (Pharmacia) with a modified polylinker; Site 1: Not I; Site 2: Eco RI; plasmid DNA from the normalized library NCI CGAP Col6 was prepared, and ss circles were made in vitro. Following HAP purification, this DNA was used as tracer in a subtractive hybridization reaction. The driver was PCR-amplified cDNAs from a pool of 5,000 clones made from the same library (cloneIDs 1057416-1061255, and 1144584-1145351).
Subtraction by Bento Soares and M. Fatima Bonaldo. "
ORIGIN
Query Match 6.1%; Score 67; DB 10; Length 492;
Best Local Similarity 100.0%; Pred. No. 2e-20;
Matches 67; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GGCCCTTGTCTCGGCCCATGGCCAGAGGGGGTCTCTGGGGCTGGGAGCTGGGGGCT 60
Db 153 GGCCCTTGTCTCGGCCCATGGCCAGAGGGGGTCTCTGGGGCTGGGAGCTGGGGGCT 212
QY 61 GTGGCCA 67
Db 213 GTGGCCA 219

Search completed: February 25, 2004, 14:35:31
Job time : 2787 secs

[illegible]

28	21	1.9	1167	3	US-09-030-607-175
29	21	1.9	1167	4	US-09-439-313-175
30	21	1.9	1167	4	US-09-352-616A-175
31	21	1.9	1167	4	US-09-232-149A-175
32	21	1.9	1167	4	US-09-159-812-175
33	21	1.9	1167	4	US-09-636-165A-175
34	21	1.9	1167	4	US-09-685-166A-175
35	21	1.9	1167	4	US-09-115-493-175
36	21	1.9	1167	4	US-09-688-483-175
37	21	1.9	1197	4	US-09-494-921-1
38	21	1.9	1197	4	US-09-494-921-28
39	21	1.9	1197	4	US-09-494-921-29
40	21	1.9	1214	4	US-09-439-313-225
41	21	1.9	1214	4	US-09-352-616A-225
42	21	1.9	1214	4	US-09-232-149A-225
43	21	1.9	1214	4	US-09-159-812-225
44	21	1.9	1214	4	US-09-636-213-225
45	21	1.9	1214	4	US-09-685-166A-225

ALIGNMENTS

```

RESULT 1
US-09-280-116-22
; Sequence 22, Application US/09280116A
; Patent NO. 6331427
; GENERAL INFORMATION:
; APPLICANT: Robison, Keith E.
; TITLE OF INVENTION: Nucleic Acid Molecules Encoding Human Protease Homologs
; FILE REFERENCE: 5800-24, 035800/176965
; CURRENT APPLICATION NUMBER: US/09/280,116A
; CURRENT FILING DATE: 1999-03-26
; NUMBER OF SEQ ID NOS: 268
; SOFTWARE: Patent in Ver. 2.0
; SEQ ID NO 22
; LENGTH: 610
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: trypsin-like serine proteases
;
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(610)
; OTHER INFORMATION: n = a, t, c, or g
;
; US-09-280-116-22

```

```

Query Match          2.2%; Score 24; DB 4; Length 610;
Best Local Similarity 100.0%; Pred. No. 0.082;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY
190 TGGCCCTTGGCAAGTGAGCCTGCAC 213

b6
102 TGGCCCTTGGCAAGTGAGCCTGCAC 125

```

RESULT 2
US-09-280-116-28/c
; Sequence 28, Application US/09280116A
; Patent No. 6331427
; GENERAL INFORMATION:
; APPLICANT: Robison, Keith E.
; TITLE OF INVENTION: Nucleic Acid Molecules
; FILE REFERENCE: 5800-24, 035800/1769
; CURRENT APPLICATION NUMBER: US/09/28
; CURRENT FILING DATE: 1999-03-26
; NUMBER OF SEQ ID NOS: 268
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 28
; LENGTH: 340
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:

```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search. using sw model

Run on: February 25, 2004, 12:20:46 ; Search time 110 Seconds
(without alignments)
5559.599 Million cell updates

Title: US-10-037-417-45

Perfect score: 1102
Sequence: 1 gggcccttctcctgggccat.....gggttctgatggggcctcc 1102

Scoring table: OLIGO_NUC
Gapop 60.0 Gapext 60.0

Searched: 682709 seqs, 277475446 residues

```
Word size : 0
Total number of bits satisfying chosen parameters: 1365418
```

```
Minimum DB seg length: 0
Maximum DB seg length: 2000000000
```

Post-processing: Listing first 45 summaries

```

Database : Issued Patents NA.*
1: /cgn2_6/ptodata/2/ina/5A COMB.seq.*
2: /cgn2_6/ptodata/2/ina/5B COMB.seq.*
3: /cgn2_6/ptodata/2/ina/6A COMB.seq.*
4: /cgn2_6/ptodata/2/ina/6B COMB.seq.*
5: /cgn2_6/ptodata/2/ina/PTCUS COMB.seq.*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq.*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	24	2.2	610	4	US-09-280-116-22	Sequence 22, Appl
2	23	2.1	340	4	US-09-280-116-28	Sequence 28, Appl
3	23	2.1	340	4	US-09-280-116-29	Sequence 29, Appl
C	22	2.0	1142	4	US-09-386-642-8	Sequence 8, Appl
4	22	2.0	1169	4	US-09-386-642-7	Sequence 7, Appl
5	22	2.0	1169	4	US-09-386-642-7	Sequence 120, App
C	6	2.1	1.9	90	US-09-020-958-120	Sequence 120, App
C	7	2.1	1.9	90	US-09-030-607-120	Sequence 120, App
C	8	2.1	1.9	90	US-09-439-313-120	Sequence 120, App
C	9	2.1	1.9	90	US-09-352-616A-120	Sequence 120, App
C	10	2.1	1.9	90	US-09-232-149A-120	Sequence 120, App
C	11	2.1	1.9	90	US-09-159-813-120	Sequence 120, App
C	12	2.1	1.9	90	US-09-636-215-120	Sequence 120, App
C	13	2.1	1.9	90	US-09-685-168A-120	Sequence 120, App
C	14	2.1	1.9	90	US-09-115-453-120	Sequence 120, App
C	15	2.1	1.9	90	US-09-688-489-120	Sequence 120, App
C	16	2.1	1.9	504	US-09-636-215-839	Sequence 839, App
17	21	1.9	504	4	US-09-685-166A-839	Sequence 839, App
18	21	1.9	700	4	US-09-280-116-68	Sequence 68, Appl
19	21	1.9	765	4	US-09-439-313-524	Sequence 524, App
20	21	1.9	765	4	US-09-636-215-524	Sequence 524, App
21	21	1.9	765	4	US-09-685-166A-524	Sequence 524, App
22	21	1.9	871	4	US-08-744-026-2	Sequence 2, Appl
23	21	1.9	871	2	US-09-102-732-2	Sequence 2, Appl
24	21	1.9	871	3	US-09-261-767-2	Sequence 2, Appl
25	21	1.9	871	4	US-08-969-987-7	Sequence 7, Appl
26	21	1.9	1037	4	US-09-386-642-60	Sequence 60, Appl
27	21	1.9	1167	3	US-09-020-956-175	Sequence 175, App

OTHER INFORMATION: trypsin-like serine proteases
US-09-280-116-28

Query Match 2.1%; Score 23; DB 4; Length 340;
Best Local Similarity 100.0%; Pred. No. 0.25;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 745 TCTGGGGGGCCCTGCTGTGA 767
Db 241 TCTGGGGGGCCCTGCTGTGA 219

RESULT 3

US-09-280-116-29/c
Sequence 29, Application US/09280116A
Patent No. 6331427
GENERAL INFORMATION:
APPLICANT: Robison, Keith E.
TITLE OF INVENTION: Nucleic Acid Molecules Encoding Human Protease Homologs
FILE REFERENCE: 5800-24, 035800/176965
CURRENT APPLICATION NUMBER: US/09/280,116A
CURRENT FILING DATE: 1999-03-26
NUMBER OF SEQ ID NOS: 268
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 29
LENGTH: 340
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: trypsin-like serine proteases
US-09-280-116-29

Query Match 2.1%; Score 23; DB 4; Length 340;
Best Local Similarity 100.0%; Pred. No. 0.25;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 745 TCTGGGGGGCCCTGCTGTGA 767
Db 241 TCTGGGGGGCCCTGCTGTGA 219

RESULT 4

US-09-386-642-8
Sequence 8, Application US/093866642
Patent No. 6420157
GENERAL INFORMATION:
APPLICANT: Darrow, Andrew
APPLICANT: Qi, Jensen
TITLE OF INVENTION: Zymogen Activation System
FILE REFERENCE: ORT-1028
CURRENT APPLICATION NUMBER: US/09/386,642
CURRENT FILING DATE: 1999-08-31
NUMBER OF SEQ ID NOS: 60
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 8
LENGTH: 1142
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: Fusion gene
OTHER INFORMATION: with homo sapien serine protease catalytic domain
US-09-386-642-8

Query Match 2.0%; Score 22; DB 4; Length 1142;
Best Local Similarity 100.0%; Pred. No. 0.69;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 731 CCTGCCAGGGTGACTCTGGGG 752
Db 704 CCTGCCAGGGTGACTCTGGGG 725

RESULT 5

US-09-386-642-7
Sequence 7, Application US/093866642
Patent No. 6420157
GENERAL INFORMATION:
APPLICANT: Darrow, Andrew
APPLICANT: Qi, Jensen
APPLICANT: Andrade-Gordon, Patricia
TITLE OF INVENTION: Zymogen Activation System
FILE REFERENCE: ORT-1028
CURRENT APPLICATION NUMBER: US/09/386,642
CURRENT FILING DATE: 1999-08-31
NUMBER OF SEQ ID NOS: 60
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 7
LENGTH: 1169
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: Fusion gene
OTHER INFORMATION: with homo sapien serine protease catalytic domain
US-09-386-642-7

Query Match 2.0%; Score 22; DB 4; Length 1169;
Best Local Similarity 100.0%; Pred. No. 0.69;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 731 CCTGCCAGGGTGACTCTGGGG 752
Db 731 CCTGCCAGGGTGACTCTGGGG 752

RESULT 6

US-09-020-956-120/c
Sequence 120, Application US/09020956
Patent No. 6261562
GENERAL INFORMATION:
APPLICANT: Xu, Jiangchun
APPLICANT: Dillin, Davin C.
TITLE OF INVENTION: COMPOUNDS FOR IMMUNOTHERAPY OF PROSTATE CANCER AND METHODS FO
NUMBER OF SEQUENCES: 178
CORRESPONDENCE ADDRESS:
ADDRESSER: SEED and BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: WA
COUNTRY: USA
ZIP: 98104
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/020,956
FILING DATE: 09-FEB-1998
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Maki, David J.
REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 210121.427C2
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 120:
SEQUENCE CHARACTERISTICS:
LENGTH: 90 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
ORIGINAL SOURCE:
ORGANISM: Homo sapiens

US-09-020-956-120
Query Match 1.9%; Score 21; DB 3; Length 90;
Best Local Similarity 100.0%; Pred. No. 2.2;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 739 GGTGACTCTGGGGGCCCTG 759
Db 34 GGTGACTCTGGGGGCCCTG 14
RESULT 7
US-09-030-607-120/c
; Sequence 120, Application US/09030607
; Patent No. 6262245
; GENERAL INFORMATION:
; APPLICANT: Xu, Jiangchun
; APPLICANT: Dillon, Davin C.
; TITLE OF INVENTION: COMPOUNDS FOR IMMUNOTHERAPY OF PROSTATE CANCER AND METHODS FO
; NUMBER OF SEQUENCES: 224
; CORRESPONDENCE ADDRESS:
; ADDRESSER: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: WA
; COUNTRY: USA
; ZIP: 98104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/030,607
; FILING DATE: 25-FEB-1998
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.427C3
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 120:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 90 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cdna
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
US-09-030-607-120
Query Match 1.9%; Score 21; DB 3; Length 90;
Best Local Similarity 100.0%; Pred. No. 2.2;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 739 GGTGACTCTGGGGGCCCTG 759
Db 34 GGTGACTCTGGGGGCCCTG 14
RESULT 8
US-09-439-313-120/c
; Sequence 120, Application US/09439313
; Patent No. 6329505
; GENERAL INFORMATION:
; APPLICANT: Xu, Jiangchun
; APPLICANT: Dillon, Davin C.
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: Harlocker, Susan Louise
; APPLICANT: Jiang Yuqui

; APPLICANT: Reed, Steven G.
; APPLICANT: Kalos, Michael
; APPLICANT: Fanger, Gary
; APPLICANT: Retter, Mark
; APPLICANT: Day, Craig
; APPLICANT: Solk, John
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THERAPY AND
; FILE REFERENCE: 210121.427C9
; CURRENT APPLICATION NUMBER: US/09/439,313
; CURRENT FILING DATE: 1999-11-12
; NUMBER OF SEQ ID NOS: 575
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 120
; LENGTH: 90
; TYPE: DNA
; ORGANISM: Homo sapien
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)...(90)
; OTHER INFORMATION: n = A,T,C or G
US-09-439-313-120
Query Match 1.9%; Score 21; DB 4; Length 90;
Best Local Similarity 100.0%; Pred. No. 2.2;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 739 GGTGACTCTGGGGGCCCTG 759
Db 34 GGTGACTCTGGGGGCCCTG 14
RESULT 9
US-09-352-616A-120/c
; Sequence 120, Application US/09352616A
; Patent No. 6395278
; GENERAL INFORMATION:
; APPLICANT: Dillon, Davin C.
; APPLICANT: Harlocker, Susan Louise
; APPLICANT: Jiang, Yuqui
; APPLICANT: Xu, Jiangchun
; APPLICANT: Mitcham, Jennifer Lynn
; TITLE OF INVENTION: COMPOUNDS FOR IMMUNOTHERAPY AND DIAGNOSIS
; FILE REFERENCE: 210121.427C8
; CURRENT APPLICATION NUMBER: US/09/352,616A
; CURRENT FILING DATE: 1999-07-13
; NUMBER OF SEQ ID NOS: 472
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 120
; LENGTH: 90
; TYPE: DNA
; ORGANISM: Homo sapien
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)...(90)
; OTHER INFORMATION: n = A,T,C or G
US-09-352-616A-120
Query Match 1.9%; Score 21; DB 4; Length 90;
Best Local Similarity 100.0%; Pred. No. 2.2;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 739 GGTGACTCTGGGGGCCCTG 759
Db 34 GGTGACTCTGGGGGCCCTG 14
RESULT 10
US-09-232-149A-120/c
; Sequence 120, Application US/09232149A
; Patent No. 6465611
; GENERAL INFORMATION:
; APPLICANT: Harlocker, Susan Louise
; APPLICANT: Jiang Yuqui

; APPLICANT: Xu, Jiangchun.
 ; APPLICANT: Dillon, Davin C.
 ; APPLICANT: Mitcham, Jennifer Lynn
 ; TITLE OF INVENTION: COMPOUNDS FOR IMMUNOTHERAPY OF PROSTATE
 ; TITLE OF INVENTION: CANCER AND METHODS FOR THEIR USE
 ; FILE REFERENCE: 210121.427C6
 ; CURRENT APPLICATION NUMBER: US/09/232,149A
 ; CURRENT FILING DATE: 1999-01-15
 ; NUMBER OF SEQ ID NOS: 338
 ; SOFTWARE: FastSeq for Windows Version 3.0
 ; SEQ ID NO 120
 ; LENGTH: 90
 ; TYPE: DNA
 ; ORGANISM: Homo sapien
 ; FEATURE:
 ; NAME/KEY: misc.feature
 ; LOCATION: (1)...(90)
 ; OTHER INFORMATION: n = A,T,C or G
 US-09-232-149A-120

Query Match 1.9%; Score 21; DB 4; Length 90;
 Best Local Similarity 100.0%; Pred. No. 2.2;
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 739 GGTGACTCTGGGGGCCCTG 759
 Db 34 GGTGACTCTGGGGGCCCTG 14

RESULT 11

US-09-159-812-120/c
 ; Sequence 120, Application US/09159812A
 ; Patent No. 6613872
 ; GENERAL INFORMATION:
 ; APPLICANT: Xu, Jiangchun
 ; APPLICANT: Dillon, Davin C.
 ; TITLE OF INVENTION: COMPOUNDS FOR IMMUNOTHERAPY OF
 ; TITLE OF INVENTION: PROSTATE CANCER AND METHODS FOR THEIR USE
 ; FILE REFERENCE: 210121.428C5
 ; CURRENT APPLICATION NUMBER: US/09/159,812A
 ; CURRENT FILING DATE: 1998-09-23
 ; NUMBER OF SEQ ID NOS: 306
 ; SOFTWARE: Fast-SEQ for Windows Version 3.0
 ; SEQ ID NO 120
 ; LENGTH: 90
 ; TYPE: DNA
 ; ORGANISM: Homo sapien
 ; FEATURE:
 ; NAME/KEY: misc.feature
 ; LOCATION: (1)...(90)
 ; OTHER INFORMATION: n = A,T,C or G
 US-09-159-812-120

Query Match 1.9%; Score 21; DB 4; Length 90;
 Best Local Similarity 100.0%; Pred. No. 2.2;
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 739 GGTGACTCTGGGGGCCCTG 759
 Db 34 GGTGACTCTGGGGGCCCTG 14

RESULT 12

US-09-636-215-120/c
 ; Sequence 120, Application US/09636215
 ; Patent No. 6620922
 ; GENERAL INFORMATION:
 ; APPLICANT: Xu, Jiangchun
 ; APPLICANT: Dillon, Davin C.
 ; APPLICANT: Mitcham, Jennifer L.
 ; APPLICANT: Harlocker, Susan L.
 ; APPLICANT: Jiang, Yuqui
 ; APPLICANT: Henderson, Robert A.

; APPLICANT: Kalos, Michael D.
 ; APPLICANT: Fanger, Gary R.
 ; APPLICANT: Retter, Marc W.
 ; APPLICANT: Stolk, John A.
 ; APPLICANT: Day, Craig H.
 ; APPLICANT: Vedvick, Thomas S.
 ; APPLICANT: Carter, Darrick
 ; APPLICANT: Li, Samuel
 ; APPLICANT: Wang, Aijun
 ; APPLICANT: Skeiky, Yasir A.W.
 ; APPLICANT: Hepler, William
 ; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
 ; TITLE OF INVENTION: DIAGNOSIS OF PROSTATE CANCER
 ; FILE REFERENCE: 210121.42717C17
 ; CURRENT APPLICATION NUMBER: US/09/636,215
 ; CURRENT FILING DATE: 2000-08-10
 ; NUMBER OF SEQ ID NOS: 852
 ; SOFTWARE: FastSeq for Windows Version 3.0
 ; SEQ ID NO 120
 ; LENGTH: 90
 ; TYPE: DNA
 ; ORGANISM: Homo sapien
 ; FEATURE:
 ; NAME/KEY: misc.feature
 ; LOCATION: (1)...(90)
 ; OTHER INFORMATION: n = A,T,C or G
 US-09-636-215-120

Query Match 1.9%; Score 21; DB 4; Length 90;
 Best Local Similarity 100.0%; Pred. No. 2.2;
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 739 GGTGACTCTGGGGGCCCTG 759
 Db 34 GGTGACTCTGGGGGCCCTG 14

RESULT 13

US-09-685-166A-120/c
 ; Sequence 120, Application US/09685166A
 ; Patent No. 6630305
 ; GENERAL INFORMATION:
 ; APPLICANT: Xu, Jiangchun
 ; APPLICANT: Dillon, Davin C.
 ; APPLICANT: Mitcham, Jennifer L.
 ; APPLICANT: Harlocker, Susan L.
 ; APPLICANT: Jiang, Yuqui
 ; APPLICANT: Henderson, Robert A.
 ; APPLICANT: Kalos, Michael D.
 ; APPLICANT: Fanger, Gary R.
 ; APPLICANT: Retter, Marc W.
 ; APPLICANT: Stolk, John A.
 ; APPLICANT: Day, Craig H.
 ; APPLICANT: Vedvick, Thomas S.
 ; APPLICANT: Carter, Darrick
 ; APPLICANT: Li, Samuel
 ; APPLICANT: Wang, Aijun
 ; APPLICANT: Skeiky, Yasir A.W.
 ; APPLICANT: Hepler, William
 ; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
 ; TITLE OF INVENTION: DIAGNOSIS OF PROSTATE CANCER
 ; FILE REFERENCE: 210121.427C21
 ; CURRENT APPLICATION NUMBER: US/09/685,166A
 ; CURRENT FILING DATE: 2000-10-10
 ; NUMBER OF SEQ ID NOS: 898
 ; SOFTWARE: Fast-SEQ for Windows Version 3.0
 ; SEQ ID NO 120
 ; LENGTH: 90
 ; TYPE: DNA
 ; ORGANISM: Homo sapien
 ; FEATURE:
 ; NAME/KEY: misc.feature
 ; LOCATION: (1)...(90)

; OTHER INFORMATION: n = A,T,C or G
 US-09-685-166A-120
 Query Match 1.9%; Score 21; DB 4; Length 90;
 Best Local Similarity 100.0%; Pred. No. 2.2;
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 739 GGTGACTCTGGGGGCCCTG 759
 Db 34 GGTGACTCTGGGGGCCCTG 14

Search completed: February 25, 2004, 14:37:33
 Job time : 111 secs

RESULT 14
 US-09-115-453-120/c
 ; Sequence 120, Application US/09115453B
 ; Patent No. 6657056
 ; GENERAL INFORMATION:
 ; APPLICANT: Xu, Jiangchun
 ; APPLICANT: Dillon, Davin C.
 ; TITLE OF INVENTION: COMPOUNDS FOR IMMUNOTHERAPY OF PROSTATE CANCER AND
 ; TITLE OF INVENTION: METHODS FOR THEIR USE
 ; FILE REFERENCE: 210121.427C4
 ; CURRENT APPLICATION NUMBER: US/09/115,453B
 ; CURRENT FILING DATE: 1998-07-14
 ; NUMBER OF SEQ ID NOS: 228
 ; SOFTWARE: FastSeq for Windows Version 3.0
 ; SEQ ID NO 120
 ; LENGTH: 90
 ; TYPE: DNA
 ; ORGANISM: Homo sapien
 ; FEATURE:
 ; NAME/KEY: misc feature
 ; LOCATION: (1)...(90)
 ; OTHER INFORMATION: n = A,T,C or G
 US-09-115-453-120

Query Match 1.9%; Score 21; DB 4; Length 90;
 Best Local Similarity 100.0%; Pred. No. 2.2;
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 739 GGTGACTCTGGGGGCCCTG 759
 Db 34 GGTGACTCTGGGGGCCCTG 14

RESULT 15
 US-09-688-489-120/c
 ; Sequence 120, Application US/09688489
 ; Patent No. 6664377
 ; GENERAL INFORMATION:
 ; APPLICANT: Xu, Jiangchun
 ; APPLICANT: Dillon, Davin C.
 ; APPLICANT: Mitcham, Jennifer Lynn
 ; TITLE OF INVENTION: COMPOUNDS FOR IMMUNOTHERAPY OF PROSTATE
 ; TITLE OF INVENTION: CANCER AND METHODS FOR THEIR USE
 ; FILE REFERENCE: 210121.427D2
 ; CURRENT APPLICATION NUMBER: US/09/688,489
 ; CURRENT FILING DATE: 2000-10-13
 ; NUMBER OF SEQ ID NOS: 338
 ; SOFTWARE: FastSeq for Windows Version 3.0
 ; SEQ ID NO 120
 ; LENGTH: 90
 ; TYPE: DNA
 ; ORGANISM: Homo sapien
 ; FEATURE:
 ; NAME/KEY: misc feature
 ; LOCATION: (1)...(90)
 ; OTHER INFORMATION: n = A,T,C or G
 US-09-688-489-120

Query Match 1.9%; Score 21; DB 4; Length 90;
 Best Local Similarity 100.0%; Pred. No. 2.2;
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;


```

; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR G
; TITLE OF INVENTION: EXPRESSION ANALYSIS TWO
; FILE REFERENCE: AEWICA-X-2
; CURRENT APPLICATION NUMBER: US/10/029,386
; CURRENT FILING DATE: 2001-12-20
; NUMBER OF SEQ ID NOS: 34288
; SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 7606
; LENGTH: 543
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO CHR16.3
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 5.1
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 5.7
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 4.3
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 4.9
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 3.3
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 3.6
; OTHER INFORMATION: SWISSPROT HIT: Q16651, EVALUE 7.00e-04
; OTHER INFORMATION: EST HUMAN HIT: AUI42128.1, EVALUE 0.00e+00
; OTHER INFORMATION: NT HIT: U33446.1, EVALUE 0.00e-00
US-10-029-386-7606

Query Match          6.1%; Score 67; DB 14; Length 543;
Best Local Similarity 100.0%; Pred. No. 5.1e-24;
Matches 67; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GGGCCCTTGTCTGGCCATGGCCAGAGGGGCTCTCTGGGGCTGGGCGAGCTGGGGGCT 60
DB 227 GGGCCCTTGTCTGGCCATGGCCAGAGGGGCTCTCTGGGGCTGGGCGAGCTGGGGGCT 168

QY 61 GTGGCCA 67
DB 167 GTGGCCA 161

RESULT 5
US-10-176-847-85
; Sequence 85, Application US/10176847
; Publication No. US20030068636A1
; GENERAL INFORMATION:
; APPLICANT: Velby, Pette Ole
; TITLE OF INVENTION: COMPOSITIONS, KITS, AND METHODS FOR
; TITLE OF INVENTION: IDENTIFICATION, ASSESSMENT, PREVENTION, AND THERAPY OF BREAST
; TITLE OF INVENTION: AND OVARIAN CANCER
; FILE REFERENCE: MEI-039
; CURRENT APPLICATION NUMBER: US/10/176,847
; CURRENT FILING DATE: 2002-06-21
; NUMBER OF SEQ ID NOS: 112
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 85
; LENGTH: 1733
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-176-847-85

Query Match          6.1%; Score 67; DB 14; Length 1733;
Best Local Similarity 100.0%; Pred. No. 4.4e-24;
Matches 67; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GGGCCCTTGTCTGGCCATGGCCAGAGGGGCTCTCTGGGGCTGGGCGAGCTGGGGGCT 60
DB 81 GGGCCCTTGTCTGGCCATGGCCAGAGGGGCTCTCTGGGGCTGGGCGAGCTGGGGGCT 140

QY 61 GTGGCCA 67
DB 141 GTGGCCA 147
```

```

RESULT 6
US-10-101-510-447
; Sequence 447, Application US/10101510
; Publication No. US20030148295A1
; GENERAL INFORMATION:
; APPLICANT: WAN, JACKSON
; APPLICANT: WANG, YIXIN
; TITLE OF INVENTION: EXPRESSION PROFILES AND METHODS OF USE
; FILE REFERENCE: 15117.0012
; CURRENT APPLICATION NUMBER: US/10/101,510
; CURRENT FILING DATE: 2002-03-20
; PRIOR APPLICATION NUMBER: 60/276,947
; NUMBER OF SEQ ID NOS: 805
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 447
; LENGTH: 3382
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-101-510-447

Query Match          6.1%; Score 67; DB 14; Length 3382;
Best Local Similarity 100.0%; Pred. No. 4e-24;
Matches 67; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GGGCCCTTGTCTGGCCATGGCCAGAGGGGCTCTCTGGGGCTGGGCGAGCTGGGGGCT 60
DB 212 GGGCCCTTGTCTGGCCATGGCCAGAGGGGCTCTCTGGGGCTGGGCGAGCTGGGGGCT 271

QY 61 GTGGCCA 67
DB 272 GTGGCCA 278

RESULT 7
US-10-042-865-29
; Sequence 29, Application US/10042865
; Publication No. US20040029216A1
; GENERAL INFORMATION:
; APPLICANT: Padigaru, Muralidhara
; APPLICANT: Li, Li
; APPLICANT: Zerhusen, Bryan D
; APPLICANT: Casman, Stacie J
; APPLICANT: Shenoy, Suresh G
; APPLICANT: Spytek, Kimberly
; APPLICANT: Zhong, Mei
; APPLICANT: Gangolli, Basha A
; APPLICANT: Burgess, Catherine E
; APPLICANT: Patturajan, Weera
; APPLICANT: Vernet, Corine A.M
; APPLICANT: Taylor, Sarah
; APPLICANT: Tchernev, Velizar T
; APPLICANT: Miller, Charles E
; APPLICANT: Guo, Xiaojia
; APPLICANT: Boldog, Ference L
; APPLICANT: Grosse, William M
; APPLICANT: Alsobrook II, John P
; APPLICANT: Gerlach, Valerie L
; APPLICANT: Edinger, Shlomit R
; APPLICANT: Rothenberg, Mark E
; APPLICANT: Ellerman, Karen
; APPLICANT: MacDougall, John
; APPLICANT: Malyankar, Uriel M
; APPLICANT: Millet, Isabelle
; APPLICANT: Peyman, John
; APPLICANT: Smithson, Glennda
; APPLICANT: Gunther, Erik
; APPLICANT: Stone, David
; TITLE OF INVENTION: Proteins, Polynucleotides Encoding Them and Methods of
; TITLE OF INVENTION: Using the Same
; FILE REFERENCE: 21402-537
; CURRENT APPLICATION NUMBER: US/10/042,865
```


; CURRENT FILING DATE: 2002-05-17
; PRIOR APPLICATION NUMBER: 60/260,417
; PRIOR FILING DATE: 2001-01-09
; PRIOR APPLICATION NUMBER: 60/260,831
; PRIOR FILING DATE: 2001-01-10
; PRIOR APPLICATION NUMBER: 60/272,338
; PRIOR FILING DATE: 2001-02-28
; PRIOR APPLICATION NUMBER: 60/274,876
; PRIOR FILING DATE: 2001-03-09
; PRIOR APPLICATION NUMBER: 60/284,704
; PRIOR FILING DATE: 2001-04-18
; NUMBER OF SEQ ID NOS: 264
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 29
; LENGTH: 1726
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-042-865-29

Query Match 4.8%; Score 53; DB 12; Length 1726;
Best Local Similarity 100.0%; Pred. No. 6.8e-17; Indels 0; Gaps 0;
Matches 53; Conservative 0; Mismatches 0;

QY 15 GGCCATGGCCAGAGGGGGTCTCTGGGGCTGGGAGCTGGGGGCTGTGGCCA 67
Db 225 GGCCATGGCCAGAGGGGGTCTCTGGGGCTGGGAGCTGGGGGCTGTGGCCA 277

RESULT 8
US-09-948-094-1
; Sequence 1, Application US/09948094
; Patent No. US20020090625A1
; GENERAL INFORMATION:
; APPLICANT: The Brigham and Women's Hospital, Inc.
; APPLICANT: Mok, Samuel
; TITLE OF INVENTION: Methods of Detecting Cancer Based on Prostatein

; FILE REFERENCE: 81994/282423
; CURRENT APPLICATION NUMBER: US/09/948,094
; CURRENT FILING DATE: 2001-09-07
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 1
; LENGTH: 1834
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (229)..(1260)
US-09-948-094-1

Query Match 4.8%; Score 53; DB 9; Length 1834;
Best Local Similarity 100.0%; Pred. No. 6.7e-17; Indels 0; Gaps 0;
Matches 53; Conservative 0; Mismatches 0;

QY 15 GGCCATGGCCAGAGGGGGTCTCTGGGGCTGGGAGCTGGGGGCTGTGGCCA 67
Db 225 GGCCATGGCCAGAGGGGGTCTCTGGGGCTGGGAGCTGGGGGCTGTGGCCA 277

RESULT 9
US-09-880-107-2214
; Sequence 2214, Application US/09880107
; Patent No. US20020142981A1
; GENERAL INFORMATION:
; APPLICANT: Horne, Darci T.
; APPLICANT: Vockley, Joseph G.
; APPLICANT: Scherf, Uwe
; APPLICANT: Gene Logic, Inc.
; TITLE OF INVENTION: Gene Expression Profiles in Liver Cancer
; FILE REFERENCE: 44921-5028-WO
; CURRENT APPLICATION NUMBER: US/09/880,107
; CURRENT FILING DATE: 2001-06-14

; PRIOR APPLICATION NUMBER: US 60/211,379
; PRIOR FILING DATE: 2000-06-14
; PRIOR APPLICATION NUMBER: US 60/237,054
; PRIOR FILING DATE: 2000-10-02
; NUMBER OF SEQ ID NOS: 3950
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2214
; LENGTH: 1834
; TYPE: DNA
; ORGANISM: Homo sapiens

; FEATURE:
; OTHER INFORMATION: Genbank Accession No. US20020142981A1 L41351
US-09-880-107-2214

Query Match 4.8%; Score 53; DB 9; Length 1834;
Best Local Similarity 100.0%; Pred. No. 6.7e-17; Indels 0; Gaps 0;
Matches 53; Conservative 0; Mismatches 0;

QY 15 GGCCATGGCCAGAGGGGGTCTCTGGGGCTGGGAGCTGGGGGCTGTGGCCA 67
Db 225 GGCCATGGCCAGAGGGGGTCTCTGGGGCTGGGAGCTGGGGGCTGTGGCCA 277

RESULT 10
US-09-967-768A-141
; Sequence 141, Application US/09967768A
; Patent No. US20020150877A1
; GENERAL INFORMATION:
; APPLICANT: Augustus, Meena

; TITLE OF INVENTION: Cancer Gene Determination and Therapeutic Screening Using Signatu
; FILE REFERENCE: 689290-72
; CURRENT APPLICATION NUMBER: US/09/967,768A
; CURRENT FILING DATE: 2001-09-28
; PRIOR APPLICATION NUMBER: US/60/236,109
; PRIOR FILING DATE: 2000-09-28
; PRIOR APPLICATION NUMBER: US/60/236,034
; PRIOR FILING DATE: 2000-09-28
; PRIOR APPLICATION NUMBER: US/60/236,111
; PRIOR FILING DATE: 2000-09-28
; NUMBER OF SEQ ID NOS: 325
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 141
; LENGTH: 1834
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-967-768A-141

Query Match 4.8%; Score 53; DB 9; Length 1834;
Best Local Similarity 100.0%; Pred. No. 6.7e-17; Indels 0; Gaps 0;
Matches 53; Conservative 0; Mismatches 0;

QY 15 GGCCATGGCCAGAGGGGGTCTCTGGGGCTGGGAGCTGGGGGCTGTGGCCA 67
Db 225 GGCCATGGCCAGAGGGGGTCTCTGGGGCTGGGAGCTGGGGGCTGTGGCCA 277

RESULT 11
US-10-097-340-261
; Sequence 261, Application US/10097340
; Publication No. US20030087250A1
; GENERAL INFORMATION:
; APPLICANT: John MONAHAN
; APPLICANT: Manjula GANNAVAPU
; APPLICANT: Sebastian HOERSCH
; APPLICANT: Shubhangi KAMATKAR
; APPLICANT: Steve G. KOVATS
; APPLICANT: Rachel E. MEYERS
; APPLICANT: Michael MORRISSEY
; APPLICANT: Peter OLANDT
; APPLICANT: Ami SEN
; APPLICANT: Peter VEIBY
; APPLICANT: Gordon B. MILLIS

APPLICANT: Robert C. BAST, Jr.
APPLICANT: Karen LU
APPLICANT: Rosemarie SCHMANDT
APPLICANT: Xumei ZHAO
APPLICANT: Karen GLATT
TITLE OF INVENTION: Nucleic Acid Molecules and Proteins For The Identification,
FILE REFERENCE: Assessment, Prevention, and Therapy of Ovarian Cancer
FILE REFERENCE: MRI-030
CURRENT FILING DATE: 2002-03-14
CURRENT APPLICATION NUMBER: US/10/097,340
PRIOR FILING DATE: 2001-03-14
PRIOR APPLICATION NUMBER: 60/276,025
PRIOR FILING DATE: 2001-03-14
PRIOR APPLICATION NUMBER: 60/325,149
PRIOR FILING DATE: 2001-09-26
PRIOR APPLICATION NUMBER: 60/276,026
PRIOR FILING DATE: 2001-03-14
PRIOR APPLICATION NUMBER: 60/324,967
PRIOR FILING DATE: 2001/09/26
PRIOR APPLICATION NUMBER: 60/311,732
PRIOR FILING DATE: 2001-08-10
PRIOR APPLICATION NUMBER: 60/325,102
PRIOR FILING DATE: 2001-09-26
PRIOR APPLICATION NUMBER: 60/323,580
PRIOR FILING DATE: 2001-09-19
NUMBER OF SEQ ID NOS: 363
SOFTWARE: FastSEQ for Windows Version 4.0
SEQ ID NO 261
LENGTH: 1834
TYPE: DNA
ORGANISM: Homo sapiens
US-10-097-340-261

Query Match 4.8%; Score 53; DB 14; Length 1834;
Best Local Similarity 100.0%; Pred. No. 6.7e-17;
Matches 53; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 15 GGCCATGCCCGCAGAGGGGCTCTGGGGCTGGGCACCTGGGGCTGTGGCCA 67
Db 225 GGCCATGCCCGCAGAGGGGCTCTGGGGCTGGGCACCTGGGGCTGTGGCCA 277

RESULT 12
US-09-922-217-931/c
Sequence 931, Application US/09922217
Patent No. US2002076414A1
GENERAL INFORMATION:
APPLICANT: Xu, Jiangchun
APPLICANT: Lodes, Michael J.
APPLICANT: Secrist, Heather
APPLICANT: Benson, Darin R.
APPLICANT: Meagher, Madeleine Joy
APPLICANT: Stolk, John A.
APPLICANT: Wang, Tongtong
APPLICANT: Jiang, Yuqiu
APPLICANT: Smith, Carole Lynn
APPLICANT: King, Gordon E.
APPLICANT: Wang, Aijun
APPLICANT: Clapper, Jonathan D.
TITLE OF INVENTION: COMPOUNDS FOR IMMUNOTHERAPY AND DIAGNOSIS
FILE REFERENCE: 210121.471C13
CURRENT FILING DATE: 2001-08-03
CURRENT APPLICATION NUMBER: US/09/922,217
NUMBER OF SEQ ID NOS: 1124
SOFTWARE: FastSEQ for Windows Version 4.0
SEQ ID NO 931
LENGTH: 596
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc_feature
LOCATION: 538_
OTHER INFORMATION: n = A,T,C or G

US-09-922-217-931

Query Match 4.6%; Score 51; DB 9; Length 596;
Best Local Similarity 100.0%; Pred. No. 8.3e-16;
Matches 51; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GGCCCTTGTCTGGCCATGGCCCGAGAGGGGCTCTGGGGCTGGGCAG 51
Db 589 GGCCCTTGTCTGGCCATGGCCCGAGAGGGGCTCTGGGGCTGGGCAG 539

RESULT 13
US-09-833-263-931/c
Sequence 931, Application US/09833263
Patent No. US20020110547A1
GENERAL INFORMATION:
APPLICANT: Wang, Aijun
APPLICANT: Clapper, Jonathan D.
APPLICANT: Stolk, John A.
APPLICANT: Meagher, Madeleine J.
TITLE OF INVENTION: COMPOUNDS FOR IMMUNOTHERAPY AND
FILE REFERENCE: 210121.471C12
CURRENT FILING DATE: 2001-04-10
CURRENT APPLICATION NUMBER: US/09/833,263
NUMBER OF SEQ ID NOS: 1093
SOFTWARE: FastSEQ for Windows Version 3.0
SEQ ID NO 931
LENGTH: 596
TYPE: DNA
ORGANISM: Homo sapien
FEATURE:
NAME/KEY: misc_feature
LOCATION: (1)_(596)
OTHER INFORMATION: n = A,T,C or G
US-09-833-263-931

Query Match 4.6%; Score 51; DB 9; Length 596;
Best Local Similarity 100.0%; Pred. No. 8.3e-16;
Matches 51; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GGCCCTTGTCTGGCCATGGCCCGAGAGGGGCTCTGGGGCTGGGCAG 51
Db 589 GGCCCTTGTCTGGCCATGGCCCGAGAGGGGCTCTGGGGCTGGGCAG 539

RESULT 14
US-10-025-380-931/c
Sequence 931, Application US/10025380
Publication No. US20020182191A1
GENERAL INFORMATION:
APPLICANT: Xu, Jiangchun
APPLICANT: Lodes, Michael J.
APPLICANT: Secrist, Heather
APPLICANT: Benson, Darin R.
APPLICANT: Meagher, Madeleine Joy
APPLICANT: Stolk, John A.
APPLICANT: Wang, Tongtong
APPLICANT: Jiang, Yuqiu
APPLICANT: Smith, Carole L.
APPLICANT: King, Gordon E.
APPLICANT: Wang, Aijun
APPLICANT: Clapper, Jonathan D.
APPLICANT: Skeiky, Yasir A. W.
APPLICANT: Fanger, Gary R.
APPLICANT: Vedvick Thomas S.
APPLICANT: Carter, Darrick
TITLE OF INVENTION: COMPOUNDS FOR IMMUNOTHERAPY AND DIAGNOSIS
FILE REFERENCE: 210121.471C14
CURRENT FILING DATE: 2001-12-19
CURRENT APPLICATION NUMBER: US/10/025,380
NUMBER OF SEQ ID NOS: 1129

! SOFTWARE: FastSeq for Windows Version 4.0
! SEQ ID NO 931
! LENGTH: 596
! TYPE: DNA
! ORGANISM: Homo sapiens
! FEATURE:
! NAME/KEY: misc_feature
! LOCATION: 538
! OTHER INFORMATION: n = A,T,C or G
US-10-025-380-931

Query Match 4.6%; Score 51; DB 13; Length 596;
Best Local Similarity 100.0%; Pred. No. 8.3e-16;
Matches 51; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GGGCCCTTCTCTGGCCCATGGCCAGAGGGGGTCTCTGGGGCTGGGGCTGGGCAG 51
DB 589 GGGCCCTTCTCTGGCCCATGGCCAGAGGGGGTCTCTGGGGCTGGGGCTGGGCAG 539

RESULT 15

US-10-042-865-31
! Sequence 31, Application US/10042865
! Publication No. US20040029216A1
! GENERAL INFORMATION:
! APPLICANT: Padigaru, Muralidhara
! APPLICANT: Li, Li
! APPLICANT: Zerhusen, Bryan D
! APPLICANT: Casman, Stacie J
! APPLICANT: Shenoy, Suresh G
! APPLICANT: Spytek, Kimberly
! APPLICANT: Zhong, Mei
! APPLICANT: Gangolli, Esha A
! APPLICANT: Burgess, Catherine E
! APPLICANT: Patturajan, Meera
! APPLICANT: Vernet, Corine A.M
! APPLICANT: Taylor, Sarah
! APPLICANT: Tchernev, Velizar T
! APPLICANT: Miller, Charles E
! APPLICANT: Guo, Xiaojia
! APPLICANT: Boldog, Renence L
! APPLICANT: Grosse, William M
! APPLICANT: Alsobrook II, John P
! APPLICANT: Gerlach, Valerie L
! APPLICANT: Edinger, Shlomit R
! APPLICANT: Rothenberg, Mark E
! APPLICANT: Ellerman, Karen
! APPLICANT: MacDougall, John
! APPLICANT: Maliyankar, Uriel M
! APPLICANT: Millet, Isabelle
! APPLICANT: Peyman, John
! APPLICANT: Smithson, Glenda
! APPLICANT: Gunther, Erik
! APPLICANT: Stone, David
! TITLE OF INVENTION: Proteins, Polynucleotides Encoding Them and Methods of
! FILE REFERENCE: 21402-537
! CURRENT APPLICATION NUMBER: US/10/042,865
! PRIOR FILING DATE: 2002-05-17
! PRIOR APPLICATION NUMBER: 60/260,417
! PRIOR FILING DATE: 2001-01-09
! PRIOR APPLICATION NUMBER: 60/260,831
! PRIOR FILING DATE: 2001-01-10
! PRIOR APPLICATION NUMBER: 60/272,338
! PRIOR FILING DATE: 2001-02-28
! PRIOR APPLICATION NUMBER: 60/274,876
! PRIOR FILING DATE: 2001-03-09
! PRIOR APPLICATION NUMBER: 60/284,704
! PRIOR FILING DATE: 2001-04-18
! NUMBER OF SEQ ID NOS: 264
! SOFTWARE: PatentIn Ver. 2.1
! SEQ ID NO 31
! LENGTH: 1161

! TYPE: DNA
! ORGANISM: Homo sapiens
US-10-042-865-31

Query Match 4.4%; Score 49; DB 12; Length 1161;
Best Local Similarity 100.0%; Pred. No. 8.1e-15;
Matches 49; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 19 ATGGCCCAAGAGGGGGTCTCTGGGGCTGGGCAGCTGGGGGCTGTGGCCA 67
DB 1 ATGGCCCAAGAGGGGGTCTCTGGGGCTGGGCAGCTGGGGGCTGTGGCCA 49

Search completed: February 25, 2004, 15:44:57
Job time : 457 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 25, 2004, 14:35:42 ; Search time 100 Seconds
(without alignments)
1008.695 Million cell updates/sec

Title: US-10-037-417-46

Perfect score: 1953
Sequence: 1 MAQKGVLPGLGAVANS...TKSLVLPWLSPLSLGLMGFP 357

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A Geneseq_29Jan04.*

- 1: geneseqp1980s.*
- 2: geneseqp1990s.*
- 3: geneseqp2000s.*
- 4: geneseqp2001s.*
- 5: geneseqp2002s.*
- 6: geneseqp2003as.*
- 7: geneseqp2003bs.*
- 8: geneseqp2004s.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1953	100.0	357	5	ABB09524 Human pro
2	1874	96.0	344	5	ABB09523 Human pro
3	1505	77.1	818	5	ABB98135 Human PMM
4	1505	77.1	818	5	AAU82753 Amino aci
5	1202	61.5	766	4	AAU93568 Human pro
6	970	49.7	178	5	AAU75082 Human pro
7	804.5	41.2	198	4	ABG04853 Novel hum
8	690	35.3	343	5	ABG96402 Human ova
9	690	35.3	343	5	AAU78547 Human pro
10	690	35.3	343	5	ABG07285 Amino aci
11	690	35.3	343	6	ABJ37067 Human bre
12	690	35.3	343	7	ADD47565 Human pro
13	690	35.3	343	7	ADD47561 Human pro
14	645	33.0	307	5	ABG98415 Human NOV
15	627.5	32.1	339	6	ABG72018 Mouse cha
16	612	31.3	386	5	ABG98416 Human NOV
17	593.5	30.4	282	5	ABP61011 Novel hum
18	590	30.2	272	5	ABG07286 Human pro
19	590	30.2	280	5	ABP61010 Novel hum
20	590	30.2	280	5	ABG91414 Primate L
21	590	30.2	284	4	AAU00467 Human ser
22	582.5	29.8	346	5	AAU74748 Human pro
23	582	29.8	691	5	AAE14348 Human pro
24	575.5	29.5	255	4	AA667514 Amino aci
25	574	29.4	262	5	ABB07284 Human pro

26	568.5	29.1	328	4	AAU41174	Aam41174 Human pol
27	567	29.0	389	5	AAU75907	Aau75907 Human epi
28	566	29.0	297	2	AAW77304	Aaw77304 Amino aci
29	563	28.8	290	3	AAU73388	Aay73388 HTRM clon
30	563	28.8	290	4	AAU12282	Aau12282 Human pro
31	563	28.8	290	4	AAU73945	Aab73945 Human pro
32	563	28.8	290	4	AAU03821	Aae03821 Human gen
33	563	28.8	290	5	ABG64545	ABG64545 Human alb
34	563	28.8	290	6	ABO17726	ABO17726 Novel hum
35	563	28.8	290	6	ABU80980	ABu80980 Human PRO
36	563	28.8	290	6	ABU66680	ABu66680 Human PRO
37	563	28.8	290	6	ABU59761	ABu59761 Novel sec
38	563	28.8	290	6	ABO24951	ABo24951 Human sec
39	563	28.8	290	6	ABG73394	ABg73394 Human ser
40	563	28.8	290	6	ABU66956	ABu66956 Human sec
41	563	28.8	290	6	ADA45741	Ada45741 Novel hum
42	563	28.8	290	6	ADA76172	Ada76172 Human PRO
43	563	28.8	290	6	ADA18822	Ada18822 Human PRO
44	563	28.8	290	6	ADA61445	Ada61445 Homo sapi
45	563	28.8	290	6	ADB19230	ADB19230 Novel hum

ALIGNMENTS

RESULT 1

ABB09524
ID ABB09524 standard; protein; 357 AA.

XX ABB09524;

XX 01-NOV-2002 (first entry)

Human prostatic precursor-like NOV14b protein, SEQ ID NO:46.

Human; NOVX; neurological disorder; Alzheimer's disease;
Huntington's disease; Parkinson's disease; pain; behavioural disorder;
addiction; tubercous sclerosis; cancer; immune disorder; allergy;
autoimmune disease; myasthenia gravis; asthma; arthritis; diabetes;
thyroiditis; cardiovascular disease; hypertension; reproductive disorder;
endometriosis; incontinence; psoriasis; scleroderma; alopecia; ulcer;
pancreatitis; cirrhosis; glomerular endotheliosis; bacterial infection;
polycystic kidney disease; endocrine disorder; obesity; cardiomyopathy;
atherosclerosis; cell signal processing-related disorder;
metabolic pathway regulation disorder; cytostatic; neuroprotective;
antiinflammatory; immunosuppressive; analgesic; antihypertensive;
dermatological; antibacterial; antiarthritic; hepatotropic; neurogenesis;
differentiation; proliferation; motility; haematopoiesis; wound healing;
angiogenesis; forensic biology; transgenic animal; drug screening;
gene therapy; NOV14b; prostatic precursor-like.

Homo sapiens.

WO200253742-A2.

11-JUL-2002.

07-JAN-2002; 2002MO-US000375.

05-JAN-2001; 2001US-0260018P.

08-JAN-2001; 2001US-0260360P.

28-FEB-2001; 2001US-0272411P.

02-MAR-2001; 2001US-0272817P.

05-JUL-2001; 2001US-0303231P.

12-JUL-2001; 2001US-0305060P.

10-SEP-2001; 2001US-0318405P.

12-SEP-2001; 2001US-0318700P.

04-JAN-2002; 2002US-00037417.

(CURA-) CURAGEN CORP.

Kekuda R, Alschbrook JP, Tchernev VT, Liu X, Spytek KA,
Patturajan M, Grosse WM, Lepley DM, Burgess CE, Vernet CAM, Li L;

PI	Gorman L, Edinger S, Sciore P, Ellerman K, Malyankar U;	301	TOPKQTQSDCLHQTAFDLSARILLRPLSHISVGVSTGKSLVLPWLSPLHGLWGPF	357
PI	Rochenberg M, Stone D, Boldog F, Guo X, Shenoy S, Anderson D;			
PI	Padigar M, Taupier R, Miller CE, Eisen A;	301	TOPKQTQSDCLHQTAFDLSARILLRPLSHISVGVSTGKSLVLPWLSPLHGLWGPF	357
XX				
DR	WPI; 2002-583619/62.			
DR	N-PSDB; ABQ93902.			
XX				
PT	Novel polypeptides and nucleic acids homologous to transmembrane			
PT	receptor, thymosin, neuromodulin-like family of proteins for diagnosing,			
PT	treating cancer, atherosclerosis, neurological, skin and autoimmune			
PT	disorders.			
XX				
PS	Claim 1c; Page 143; 323pp; English.			
XX				
CC	The invention relates to 24 novel human proteins designated NOV1-NOV14			
CC	(ABB09501-ABB09524), collectively referred to as NOVX proteins, and			
CC	nucleic acids encoding them (ABQ93879-ABQ93902). NOVX proteins and			
CC	nucleotides are useful in the treatment, diagnosis or prevention of NOVX-			
CC	associated disorders or in the manufacture of a medicament for treating			
CC	such disorders, with specific applications described for each of the 24			
CC	NOVX proteins, based on their homology to known proteins. Various			
CC	disorders are associated with NOVX proteins including neurological			
CC	disorders (e.g., Alzheimer's, Huntington's and Parkinson's diseases),			
CC	pain, behavioural disorders, addiction, tuberculous sclerosis, cancers			
CC	(e.g., colorectal cancer, leukaemia and osteosarcoma), immune disorders			
CC	(e.g., allergies and autoimmune diseases), myasthenia gravis, asthma,			
CC	various forms of arthritis, diabetes, thyroiditis, cardiovascular disease			
CC	(e.g., hypertension), reproductive disorders, endometriosis,			
CC	incontinence, psoriasis, scleroderma, alopecia, ulcers, pancreatitis,			
CC	cirrhosis, glomerular endotheliosis, polycystic kidney disease, endocrine			
CC	disorders, obesity, bacterial infections and particularly cardiomyopathy,			
CC	atherosclerosis, cell signal processing-related disorders and disorders			
CC	of metabolic pathway regulation. NOVX nucleic acids and polypeptides may			
CC	be used to identify cellular receptors or downstream effectors which			
CC	binds to a NOVX protein, and are also useful as targets for the			
CC	identification of small molecules that modulate or inhibit processes such			
CC	as neurogenesis, cell differentiation, cell motility, cellular			
CC	proliferation, haematopoiesis, wound healing and angiogenesis. NOVX			
CC	nucleic acid sequences can be used to identify a cell or tissue type and			
CC	are useful as a source of primers or probes for forensic biology and for			
CC	identifying and cloning NOVX homologues in other cell types. Cells			
CC	comprising NOVX nucleic acids are useful for producing non-human			
CC	transgenic animals which are useful for studying the function and			
CC	activity of NOVX proteins and for identifying and evaluating modulators			
CC	of NOVX activity. The present sequence represents the prostatin precursor			
XX	-like protein NOV14b			
SQ	Sequence 357 AA;			
	Query Match	100.0%;	Score 1953;	DB 5; Length 357;
	Best Local Similarity	100.0%;	Pred. No. 5.2e-146;	
	Matches 357; Conservative	0;	Mismatches	0; Indels
				0; Gaps
QY	1	MAQKVLGPGQLGAVANDSDSLYLVSFGPARGPPYCGRPEPSARIYVGGNAQGTWPM	60	
DB	1	MAQKVLGPGQLGAVANDSDSLYLVSFGPARGPPYCGRPEPSARIYVGGNAQGTWPM	60	
QY	61	QVSLHHGGHICGGSLIAPSWLSAAHCFMTWNTGTEPAEWSVLGVHSQDGLDGAHTR	120	
DB	61	QVSLHHGGHICGGSLIAPSWLSAAHCFMTWNTGTEPAEWSVLGVHSQDGLDGAHTR	120	
QY	121	AVAAIIVVPANYTSQVELGADLALLRLASPASLGPVAVPVCPLPRASHRFVHGTCATWATGMD	180	
DB	121	AVAAIIVVPANYTSQVELGADLALLRLASPASLGPVAVPVCPLPRASHRFVHGTCATWATGMD	180	
QY	181	VOEADPLPLPWLTQVEVLLGELATCCLYSQPGFNLTLQLPGMLCAGYPEGRRDTCQ	240	
DB	181	VOEADPLPLPWLTQVEVLLGELATCCLYSQPGFNLTLQLPGMLCAGYPEGRRDTCQ	240	
QY	241	GDSGGPLVCEGGRWFQAGITSGFCGGRNPGVFTAVATYEAWIREQVMGSEFGPAPP	300	
DB	241	GDSGGPLVCEGGRWFQAGITSGFCGGRNPGVFTAVATYEAWIREQVMGSEFGPAPP	300	

QY	301	TOPKQTQSDCLHQTAFDLSARILLRPLSHISVGVSTGKSLVLPWLSPLHGLWGPF	357
DB	301	TOPKQTQSDCLHQTAFDLSARILLRPLSHISVGVSTGKSLVLPWLSPLHGLWGPF	357
XX			
RESULT 2			
ABB09523			
ID	ABB09523	standard; protein; 344 AA.	
XX			
AC	ABB09523;		
XX			
DT	01-NOV-2002	(first entry)	
XX			
DE	Human prostatin precursor-like NOV14a protein, SEQ ID NO:44.		
XX			
KW	Human; NOVX; neurological disorder; Alzheimer's disease;		
KW	Huntington's disease; Parkinson's disease; pain; behavioural disorder;		
KW	addiction; tuberculous sclerosis; cancer; immune disorder; allergy;		
KW	autoimmune disease; myasthenia gravis; asthma; arthritis; diabetes;		
KW	thyroiditis; cardiovascular disease; hypertension; reproductive disorder;		
KW	endometriosis; incontinence; psoriasis; scleroderma; alopecia; ulcer;		
KW	pancreatitis; cirrhosis; glomerular endotheliosis; bacterial infection;		
KW	polycystic kidney disease; endocrine disorder; obesity; cardiomyopathy;		
KW	atherosclerosis; cell signal processing-related disorder;		
KW	metabolic pathway regulation disorder; cytostatic; neuroprotective;		
KW	antiinflammatory; immunosuppressive; analgesic; antithrombotic;		
KW	dermatological; antibacterial; antiarthritic; hepatotropic; neurogenesis;		
KW	differentiation; proliferation; motility; haematopoiesis; wound healing;		
KW	angiogenesis; forensic biology; transgenic animal; drug screening;		
KW	gene therapy; NOV14a; prostatin precursor-like; chromosome 16.		
OS	Homo sapiens.		
XX			
PN	W0200253742-A2.		
XX			
PD	11-JUL-2002.		
XX			
PF	07-JAN-2002; 2002WO-US0000375.		
XX			
PR	05-JAN-2001; 2001US-0260018P.		
PR	08-JAN-2001; 2001US-0260360P.		
PR	28-FEB-2001; 2001US-0272411P.		
PR	02-MAR-2001; 2001US-0272817P.		
PR	05-JUL-2001; 2001US-0302311P.		
PR	12-JUL-2001; 2001US-0305060P.		
PR	10-SEP-2001; 2001US-0318405P.		
PR	12-SEP-2001; 2001US-0318700P.		
PR	04-JAN-2002; 2002US-00037417.		
XX			
FA	(CURA-) CURAGEN CORP.		
XX			
PI	Kekuda R, Alsobrook JP, Tcherven VT, Liu X, Spytek KA;		
PI	Patturajan M, Grose WM, Lepley DM, Burgess CE, Vernet CAM, Li L;		
PI	Gorman L, Edinger S, Sciore P, Ellerman K, Malyankar U;		
PI	Rochenberg M, Stone D, Boldog F, Guo X, Shenoy S, Anderson D;		
PI	Padigar M, Taupier R, Miller CE, Eisen A;		
XX			
DR	WPI; 2002-583619/62.		
DR	N-PSDB; ABQ93901.		
XX			
PT	Novel polypeptides and nucleic acids homologous to transmembrane		
PT	receptor, thymosin, neuromodulin-like family of proteins for diagnosing,		
PT	treating cancer, atherosclerosis, neurological, skin and autoimmune		
PT	disorders.		
XX			
PS	Claim 1c; Page 142; 323pp; English.		
XX			
CC	The invention relates to 24 novel human proteins designated NOV1-NOV14		
CC	(ABB09501-ABB09524), collectively referred to as NOVX proteins, and		
CC	nucleic acids encoding them (ABQ93879-ABQ93902). NOVX proteins and		
CC	nucleotides are useful in the treatment, diagnosis or prevention of NOVX-		
CC	associated disorders or in the manufacture of a medicament for treating		
CC	such disorders, with specific applications described for each of the 24		

CC NOVX proteins, based on their homology to known proteins. Various
CC disorders are associated with NOVX proteins including neurological
CC disorders (e.g., Alzheimer's, Huntington's and Parkinson's diseases),
CC pain, behavioural disorders, addiction, tuberosclerosis, cancers
CC (e.g., colorectal cancer, leukaemia and osteosarcoma), immune disorders
CC (e.g., allergies and autoimmune diseases), myasthenia gravis, asthma,
CC various forms of arthritis, diabetes, thyroiditis, cardiovascular disease
CC (e.g., hypertension), reproductive disorders, endometriosis,
CC incontinence, psoriasis, scleroderma, alopecia, ulcers, pancreatitis,
CC cirrhosis, glomerular endotheliosis, polycystic kidney disease, endocrine
CC disorders, obesity, bacterial infections and particularly cardiomyopathy,
CC atherosclerosis, cell signal processing-related disorders and disorders
CC of metabolic pathway regulation. NOVX nucleic acids and polypeptides may
CC be used to identify cellular receptors or downstream effectors which
CC binds to a NOVX protein, and are also useful as targets for the
CC identification of small molecules that modulate or inhibit processes such
CC as neurogenesis, cell differentiation, cell motility, cellular
CC proliferation, haematopoiesis, wound healing and angiogenesis. NOVX
CC nucleic acid sequences can be used to identify a cell or tissue type and
CC are useful as a source of primers or probes for forensic biology and for
CC identifying and cloning NOVX homologues in other cell types. Cells
CC comprising NOVX nucleic acids are useful for producing non-human
CC transgenic animals which are useful for studying the function and
CC activity of NOVX proteins and for identifying and evaluating modulators
CC of NOVX activity. The present sequence represents the prostatic precursor
CC -like protein NOV14a. The gene encoding NOV14a is located on chromosome
CC 16
XX
SQ Sequence 344 AA;

Query Match 96.0%; Score 1874; DB 5; Length 344;
Best Local Similarity 100.0%; Pred. No. 8.6e-140;
Matches 344; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MAQKGLVPGQLGAVANSYSYLGIVPSGARGPPYCGRPSPARIVGGSNAQPGTW 60
Db 1 MAQKGLVPGQLGAVANSYSYLGIVPSGARGPPYCGRPSPARIVGGSNAQPGTW 60
QY 61 QVSLHGGGHI CGSLIAPSWSLSAAHCFWNTGTLPEAPAEWSVLLGVHSQDGLDGAHTR 120
Db 61 QVSLHGGGHI CGSLIAPSWSLSAAHCFWNTGTLPEAPAEWSVLLGVHSQDGLDGAHTR 120
QY 121 AVAAIVVPANYSQVELGADLALLRIASPSASLGAVVPCVLPASHRFVHGTCWATGWD 180
Db 121 AVAAIVVPANYSQVELGADLALLRIASPSASLGAVVPCVLPASHRFVHGTCWATGWD 180
QY 181 VQEAADPLPFWLQVVELLGEATCCLYSQSGPENLTLQLPGMLCAGYEGREDTCQ 240
Db 181 VQEAADPLPFWLQVVELLGEATCCLYSQSGPENLTLQLPGMLCAGYEGREDTCQ 240
QY 241 GDSGGLVCEEGGRWPFQAGITSGFCGRRNRPVFTAVATYEAMIREQVMGSEPGAPFP 300
Db 241 GDSGGLVCEEGGRWPFQAGITSGFCGRRNRPVFTAVATYEAMIREQVMGSEPGAPFP 300
QY 301 TOPQKTSCLHQTAFDLSARILLRPLSHSVGVSTGTSLVLP 344
Db 301 TOPQKTSCLHQTAFDLSARILLRPLSHSVGVSTGTSLVLP 344

RESULT 3
ABB98135
ID ABB98135 standard; protein; 818 AA.
XX
AC ABB98135;
XX
DT 17-OCT-2002 (first entry)
XX
DE Human PMMM Incyte ID 2751509CD1.
XX
KW Human; PMMM; protein modification and maintenance molecule;
KW anticonvulsant; neuroprotective; nootropic; cytostatic; antipsoriatic;
KW antiasthmatic; dermatological; antidiabetic; antiparkinsonian;
KW antianaemic; antiinflammatory; antiulcer; antiangular; cardiant;

KW hepatotropic; osteopathic; antiemetic; antipyretic; virucide;
KW antibacterial; fungicide; gastrointestinal; antiarrhythmic; laxative;
KW haemostatic; anti-HIV; thrombolytic; anticoagulant; gynaecological;
KW cardiovascular; antiarteriosclerotic; hypotensive; vasotropic;
KW antitumour; antihemematic; immunosuppressive; antiallergic; antihypertensive;
KW nephrotropic; antitumor; thyromimetic; antiarthritic; uropathic;
KW ophthalmological; antiparasitic; tranquiliser; vulnerary; keratolytic;
KW auditory; antiseborrheic; antidepressant; neuroleptic; antiinfertility;
KW antelmintic; protozoacide; Crohn's disease; hypertension; autoimmune;
KW inflammatory; anaemia; cell proliferative; development; epithelial;
KW scabies; neurological; Alzheimer's disease; reproductive;
KW ectopic pregnancy; gene therapy; vaccine; disorder; prostasin.
XX Homo sapiens.
OS WO200246383-A2.
PN 13-JUN-2002.
PD
XX 05-DEC-2001; 2001WO-US046964.
XX
PF 08-DEC-2000; 2000US-0254399P.
PR 21-DEC-2000; 2000US-0257803P.
PR 05-JAN-2001; 2001US-0260110P.
PR 19-JAN-2001; 2001US-0262851P.
PR 25-JAN-2001; 2001US-0264623P.
XX
XX (INCY-) INCYTE GENOMICS INC.
XX
XX Yue H, Azimzai Y, Kallick DA, Baughn MR, Griffin JA, Swarnakar A;
PI Lal PG, Wallia NK, Hafalia AJA, Gandhi AR, Au-Young J, Elliott VS;
PI Ramkumar J, Thangavelu K, Lu Y, Warren BA, Lu DAM, Lee BA;
PI Tribouley CM, Arvizu C, Delegeane AM, Yao MG, Khan FA;
PI Sanjanwala MM;
XX
XX WPI; 2002-519664/55.
DR N-PSDB; ABQ75956.
XX
XX New isolated Protein Modification and Maintenance polypeptides, useful
XX for diagnosis, and treatment of e.g. gastrointestinal disorders.
PS Claim 1 (a); Page 174-176; 200pp; English.
XX
XX The invention relates to an isolated Protein Modification and Maintenance
XX (PMM) polypeptide. Polypeptides of the invention may be used in the
XX diagnosis, treatment and prevention of disorders associated with
XX decreased expression or activity of PMM. These include gastrointestinal
XX disorders (e.g. Crohn's disease), cardiovascular disorders (e.g. aneamia), cell
XX hypertension), autoimmune/inflammatory disorders (e.g. Alzheimer's disease)
XX proliferative disorders, developmental disorders, epithelial disorders
XX (e.g. scabies), neurological disorders (e.g. Alzheimer's disease)
XX reproductive disorders (e.g. ectopic pregnancy), and in gene therapy or a
XX vaccine for such diseases. They may also be used in the assessment of the
XX effects of exogenous compound on the expression of nucleic acid and amino
XX acid sequences of protein modification and maintenance molecules. The
XX current sequence represents a human PMM of the invention, which has been
XX found to have homology with rat prostasin
SQ Sequence 818 AA;

Query Match 77.1%; Score 1505; DB 5; Length 818;
Best Local Similarity 100.0%; Pred. No. 3.2e-110;
Matches 272; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 38 CGRPEPSARI VGGSSNAQPGTWQVSLHHGGHICGSLIAPSWVLAAHCFWNTGTLPE 97
Db 38 CGRPEPSARI VGGSSNAQPGTWQVSLHHGGHICGSLIAPSWVLAAHCFWNTGTLPE 97
QY 98 AAQWSVLLGVHSQDGLDGAHTRAAVAVVVPANYSQVELGADLALLRIASPSASLGPAVWP 157
Db 98 AAQWSVLLGVHSQDGLDGAHTRAAVAVVVPANYSQVELGADLALLRIASPSASLGPAVWP 157
QY 158 VCLPRASHRFVHGTCWATGWDVQEAADPLPFWLQVVELLGEATCCQLYSQGGFPN 217

KW food supplement; medical imaging; diagnostic; genetic disorder.

OS Homo sapiens.

PN WO200175067-A2.

XX 11-OCT-2001.

XX 30-MAR-2001; 2001WO-US008631.

PR 31-MAR-2000; 2000US-00540217.

PR 23-AUG-2000; 2000US-00649167.

XX (HYSE-) HYSEQ INC.

XX Drmanac RT, Liu C, Tang YT;

XX WPI; 2001-639362/73.

DR N-PSDB; AAS69040.

XX New isolated polynucleotide and encoded polypeptides, useful in diagnostics, forensics, gene mapping, identification of mutations responsible for genetic disorders or other traits and to assess biodiversity.

PS Claim 20; SEQ ID NO 35212; 103pp; English.

XX The invention relates to isolated polynucleotide (I) and polypeptide (II) sequences. (I) is useful as hybridisation probes, polymerase chain reaction (PCR) primers, oligomers, and for chromosome and gene mapping, and in recombinant production of (II). The polynucleotides are also used in diagnostics as expressed sequence tags for identifying expressed genes. (I) is useful in gene therapy techniques to restore normal activity of (II) or to treat disease states involving (II). (II) is useful for generating antibodies against it, detecting or quantitating a polypeptide in tissue, as molecular weight markers and as a food supplement. (II) and its binding partners are useful in medical imaging of sites expressing (II). (I) and (II) are useful for treating disorders involving aberrant protein expression or biological activity. The polypeptide and polynucleotide sequences have applications in diagnostics, forensics, gene mapping, identification of mutations responsible for genetic disorders or other traits to assess biodiversity and to produce other types of data and products dependent on DNA and amino acid sequences. ABG0010-ABG30377 represent novel human diagnostic amino acid sequences of the invention. Note: The sequence data for this patent did not appear in the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences

XX Sequence 198 AA;

Query Match 41.2%; Score 804.5; DB 4; Length 198;
Best Local Similarity 96.2%; Pred. No. 1.3e-55;
Matches 150; Conservative 1; Mismatches 4; Indels 1; Gaps 1;

QY 93 GTLEPAEWSVLGVHSGQDPLDGAHRAVAIVPANTYSQVELGADIALRLASPASIG 152

Db 1 GTLEPAEWSVLGVHSGQDPLDGAHRAVAIVPANTYSQVELGADIALRLASPASIG 60

QY 153 PAVMPVCLPRASHRFVHGTACWATGWDVQEDPLPWLQVELRLGEATCCCLYSQ 212

Db 61 PAVMPVCLPRASHRFVHGTACWATGWDVQEDPLPWLQVELRLGEATCCCLYSQ 120

QY 213 PGPFNLTLQLFGMLCAGYPEGRRDTCQ-GDSGGPL 247

Db 121 PGPFNLTLQLFGMLCAGYPEGRRDTCQSPREGPM 156

RESULT 8

ABG96402

ID ABG96402 standard; protein; 343 AA.

XX ABG96402;

AC ABG96402;

XX

DT 11-DEC-2002 (first entry)

XX Human ovarian cancer marker OV80.

DE Human; ovarian cancer; marker; cancer; familial history; brain disorder;

KW central nervous system disorder; bacterial meningitis; viral meningitis;
KW Alzheimer's disease; Parkinson's disease; cerebral oedema; hydrocephalus;
KW brain herniation; inflammation; encephalitis; testicular disorder;
KW nontuberculous granulomatous orchitis; connective tissue disorder;
KW heart disorder; ischaemic heart disease; atherosclerosis; neoplasm;
KW histological type; carcinogenic; ovarian cancer marker.

XX Homo sapiens.

XX WO200271928-A2.

XX 19-SEP-2002.

XX 14-MAR-2002; 2002WO-US007826.

XX 14-MAR-2001; 2001US-0276025P.

PR 14-MAR-2001; 2001US-0276036P.

PR 10-AUG-2001; 2001US-0311732P.

PR 19-SEP-2001; 2001US-0323580P.

PR 26-SEP-2001; 2001US-0324967P.

PR 26-SEP-2001; 2001US-0325102P.

PR 26-SEP-2001; 2001US-0325149P.

XX (MILL-) MILLENNIUM PHARM INC.

XX Monahan JE, Gannavarapu M, Hoersch S, Kamatkar S, Kovatis SG;

PI Meyers RE, Morrissy MP, Olandt PJ, Sen A, Vieby PO, Mills GB;

PI Bast RC, Lu K, Schmandt RE, Zhao X, Glatt K;

XX WPI; 2002-723277/78.

DR N-PSDB; ABS76501.

XX Assessing whether a patient is afflicted with ovarian cancer, useful in assessing the stage or progression of the disease, comprises comparing the expression level of a cancer marker in a sample from a patient and from a non cancer patient.

PS Disclosure; Page 385; 481pp; English.

XX The present invention relates to a new method for assessing whether a patient is afflicted with ovarian cancer. The method involves comparing the expression level of a marker in a patient sample and the normal level of expression of the marker in a control non-ovarian cancer sample, where the marker is selected from 363 cancer markers described in the specification. The method of the invention is useful in diagnosing or characterising cancer, in detecting the presence of cancer as early as possible, and the recurrence of ovarian cancer. The method may also be of particular use with patients having an enhanced risk of developing ovarian cancer (e.g. patients having a familial history of ovarian cancer). The cancer markers may be used in the management and treatment of e.g. brain and central nervous system disorders (e.g. bacterial and viral meningitis, Alzheimer's disease or Parkinson's disease), brain disorders (e.g. cerebral oedema, hydrocephalus or brain herniations), inflammations (e.g. bacterial or viral meningitis or encephalitis), testicular disorders (e.g. nontuberculous granulomatous orchitis), connective tissue disorders, or heart disorders (e.g. ischaemic heart disease or atherosclerosis). The compositions and methods may also be used in assessing the histological type of neoplasm associated with ovarian cancer, monitoring the progression of ovarian cancer, determining whether ovarian cancer has metastasized or is likely to metastasize, selecting a composition for inhibiting ovarian cancer, assessing the ovarian carcinogenic potential of a compound, or inhibiting ovarian cancer or at risk of developing ovarian cancer. The present amino acid sequence represents one of the ovarian cancer markers described in the invention

XX Sequence 343 AA;

SQ

Query Match 35.3%; Score 690; DB 5; Length 343;
 Best Local Similarity 43.6%; Pred. No. 2.7e-46;
 Matches 156; Conservative 49; Mismatches 129; Indels 24; Gaps 10;
 QY 1 MAQKGVLPQGLGAVANSYSYLY-GLVPSGP-----ARGPPYCGRPEPSARIVGGSNAQP 55
 DB 1 MAQKGVLPQGLGAVA-----ILLYLGLLRSGTGAGAEAP--CG-VAPQARITGSSAVA 53
 QY 56 GTWPMQVSLHHGGHICGGLSLIAPSWLSAAHCFMNTGTLPEPAEWSVLLGVHSQDGLD 115
 DB 54 GQMPQVSYTYEGVHVCGSLVSEQWLSAAHCFPSEHKE---AYEVKLGHAHQDLSYSE 110
 QY 116 GAHTRAVAAIVPANYSQVELGADLALLRLASPASLGPVAVPCLPRASHRFVHGTA 175
 DB 111 DAKVSTLKDIIPHSYLOEQSGDIALQLSRITFSYRIRPCLPAANASFPNGLHCTV 170
 QY 176 TCGVDVQVADPLPLPWVQLQEVLELRLGEATCCCLYSQPGPFNLTQLPGLMCA 235
 DB 171 TCGHVPASVSLLTTPKPLQQLQEVLELRLGEATCCCLYSQPGPFNLTQLPGLMCA 230
 QY 236 RDTCCQDGGGLVCEBGGFWFQAGITSGFGCGRRNRPGVFTAVATYEAIRQVWGSEP 295
 DB 231 KDACQDGGGLVCEBGGFWFQAGITSGFGCGRRNRPGVFTAVATYEAIRQVWGSEP 288
 QY 296 GFAPPTQPKTQSD---CLHQTAFIDS-ARILLRPLSHISVGVSTGKSLVLPWLS 349
 DB 289 QPRVVPQTQESQPDNLGSLHAFSSAPAGLLRPILFLPLGLALG---LLSPWLSEH 343

RESULT 9
 AAU78547
 ID AAU78547 standard; protein; 343 AA.
 AC AAU78547;
 XX
 XX
 DT 18-JUN-2002 (first entry)
 DE Human prostatic protein.
 KW Prostatic; human; malignant; cancer; ovarian cancer; breast cancer;
 KW prostate cancer; lung cancer; colon cancer.
 OS Homo sapiens.
 XX
 XX
 PN W0200221133-A2.
 XX
 PD 14-MAR-2002.
 XX
 PF 07-SEP-2001; 2001WO-US027718.
 XX
 XX
 PR 07-SEP-2000; 2000US-0231166P.
 XX
 XX
 PA (BGHM) BRIGHAM & WOMENS HOSPITAL INC.
 XX
 PI Mok SC, Wong K;
 XX
 XX
 DR WPI; 2002-292285/33.
 DR N-PSDB; ABK12241.
 XX
 XX
 PT Diagnosing cancer or susceptibility to it, useful particularly for
 PT ovarian cancer, comprises detecting increased levels of prostatic in
 PT blood or tissue.
 XX
 XX
 PS Disclosure; Page 34-36; 36pp; English.
 XX
 CC This invention relates to a novel method for determining if a human has,
 CC or is likely to develop, a malignant growth. The method comprises
 CC measuring the concentration of prostatic protein or mRNA, in a sample
 CC and comparing the result with one or more controls, where a level of
 CC prostatic expression that is significantly higher than in controls is
 CC indicative of cancer, or increased susceptibility. The assay is
 CC specifically used to detect ovarian cancer or a predisposition to it, but

CC may also be applicable to breast, prostatic, lung and colonic cancers.
 CC Using the method of the invention it was shown that in ovarian cancer,
 CC the highest levels of prostatic protein were found in stage II disease,
 CC suggesting that this marker is suitable for early detection. The present
 CC sequence represents the human prostatic protein used in the method of the
 CC invention as a marker for early detection of cancer or susceptibility to
 CC cancer
 XX
 SQ Sequence 343 AA;
 Query Match 35.3%; Score 690; DB 5; Length 343;
 Best Local Similarity 43.6%; Pred. No. 2.7e-46;
 Matches 156; Conservative 49; Mismatches 129; Indels 24; Gaps 10;
 QY 1 MAQKGVLPQGLGAVANSYSYLY-GLVPSGP-----ARGPPYCGRPEPSARIVGGSNAQP 55
 DB 1 MAQKGVLPQGLGAVA-----ILLYLGLLRSGTGAGAEAP--CG-VAPQARITGSSAVA 53
 QY 56 GTWPMQVSLHHGGHICGGLSLIAPSWLSAAHCFMNTGTLPEPAEWSVLLGVHSQDGLD 115
 DB 54 GQMPQVSYTYEGVHVCGSLVSEQWLSAAHCFPSEHKE---AYEVKLGHAHQDLSYSE 110
 QY 116 GAHTRAVAAIVPANYSQVELGADLALLRLASPASLGPVAVPCLPRASHRFVHGTA 175
 DB 111 DAKVSTLKDIIPHSYLOEQSGDIALQLSRITFSYRIRPCLPAANASFPNGLHCTV 170
 QY 176 TCGVDVQVADPLPLPWVQLQEVLELRLGEATCCCLYSQPGPFNLTQLPGLMCA 235
 DB 171 TCGHVPASVSLLTTPKPLQQLQEVLELRLGEATCCCLYSQPGPFNLTQLPGLMCA 230
 QY 236 RDTCCQDGGGLVCEBGGFWFQAGITSGFGCGRRNRPGVFTAVATYEAIRQVWGSEP 295
 DB 231 KDACQDGGGLVCEBGGFWFQAGITSGFGCGRRNRPGVFTAVATYEAIRQVWGSEP 288
 QY 296 GFAPPTQPKTQSD---CLHQTAFIDS-ARILLRPLSHISVGVSTGKSLVLPWLS 349
 DB 289 QPRVVPQTQESQPDNLGSLHAFSSAPAGLLRPILFLPLGLALG---LLSPWLSEH 343
 RESULT 10
 ABB07285
 ID ABB07285 standard; protein; 343 AA.
 AC ABB07285;
 XX
 XX
 DT 26-MAR-2002 (first entry)
 XX
 XX
 DE Amino acid sequence of protein seq Id No. 3 (Swiss Prot No. Q16651).
 KW Prostatic-like enzyme; human; prostatic-like serine protease; cytosolic;
 KW antiatherosclerotic; virucide; osteopathic; antiinflammatory; vasotropic;
 KW neuroprotective; gene therapy; antisense therapy.
 OS Homo sapiens.
 XX
 XX
 PN W0200198466-A2.
 XX
 PD 27-DEC-2001.
 XX
 PF 22-JUN-2001; 2001WO-EP007116.
 XX
 XX
 PR 23-JUN-2000; 2000US-0213474P.
 PR 22-MAR-2001; 2001US-0277612P.
 XX
 XX
 PA (FARB) BAYER AG.
 XX
 XX
 PI Xiao Y;
 XX
 DR WPI; 2002-114575/15.
 XX
 PT Novel human prostatic-like enzyme polypeptide and polynucleotide which
 PT can be regulated for treating metastasis of malignant cells,
 PT inflammation, atherosclerosis, neurodegenerative disease and pathogenic

XX PN WO2003016475-A2.
XX PD 27-FEB-2003.
XX PF 14-AUG-2002; 2002WO-US025765.
XX PR 14-AUG-2001; 2001US-0312147P.
XX PR 01-NOV-2001; 2001US-0346382P.
XX PR 26-NOV-2001; 2001US-0333347P.
XX PA (GHEO) GEN HOSPITAL CORP.
XX PA (FARB) BAYER AG.
XX PI Woolf C, D'urso D, Befort K, Costigan M;
XX DR WPI; 2003-268312/26.
XX DR GENBANK; Q16651.
XX PT New composition comprising two or more isolated polypeptides, useful for
XX PT preparing a medicament for treating pain in an animal.
XX PS Claim 1; Page; 1017pp; English.
XX CC The invention discloses a composition comprising two or more isolated rat
XX CC or human polynucleotides or a polynucleotide which represents a fragment,
XX CC derivative or allelic variation of the nucleic acid sequence. Also
XX CC claimed are a vector comprising the novel polynucleotide, a host cell
XX CC comprising the vector, a method for identifying a nucleotide sequence
XX CC which is differentially regulated in an animal subjected to pain and a
XX CC kit to perform the method, an array, a method for identifying an agent
XX CC that increases or decreases the expression of the polynucleotide sequence
XX CC that is differentially expressed in neuronal tissue of a first animal
XX CC subjected to pain, a method for identifying a compound which regulates
XX CC the expression of a polynucleotide sequence which is differentially
XX CC expressed in an animal subjected to pain, a method for identifying a
XX CC compound that regulates the activity of one or more of the
XX CC polynucleotides, a method for producing a pharmaceutical composition, a
XX CC method for identifying a compound or small molecule that regulates the
XX CC activity in an animal of one or more of the polypeptides given in the
XX CC specification, a method for identifying a compound useful in treating
XX CC pain and a pharmaceutical composition comprising the one or more
XX CC polypeptides or their antibodies. The polynucleotide or the compound that
XX CC modulates its activity is useful for preparing a medicament for treating
XX CC pain (e.g. spinal segmental nerve injury (Chung), chronic constriction
XX CC injury (CCI) and spared nerve injury (SNI)) in an animal (e.g. gene
XX CC therapy). The sequence presented is a human protein (shown in Table 2 of
XX CC the specification) which is differentially expressed during pain. Note:
XX CC The sequence data for this patent did not form part of the printed
XX CC ftp.wipo.int/pub/published_pct_sequences.

Query Match 35.3%; Score 690; DB 7; Length 343;
Best Local Similarity 43.6%; Pred. No. 2.7e-46;
Matches 156; Conservative 49; Mismatches 129; Indels 24; Gaps 10;

QY 1 MAQKGVLPQGLGAVANSYSLY GLVPSGP-----ARGPPYCGRPEPSARIYVCGSNAPQ 55
DB 1 MAQKGVLPQGLGAVA-----ILLYGLLRSGTGAEGARAP--CG-VAPQARITGSSAVA 53
QY 56 GTWPMVSLHHGGCHTCGSLIAPSVLSAAHCFMTNGTLEPAEASVLLGVHRSQDGLD 115
DB 54 GQMPWQVSYTYEGHVHVGSGLSVSEQWVLSNAHCFPSHHKE---AEVKLGAHQLDYSYE 110
QY 116 GAHTRAVAIVPANYQVELGADLALLRLASPASLGPVAVPVLPRASHRFVHGTAWA 175
DB 111 DAKVSTLKDIIPHSYVLEQSGQDIALQLSRITPSTRVIRPICLPAANASFPNGLHCTV 170
QY 176 TGHGDVQADPLPLPWLQVEVRLILGEATCQCLYSQPGFNLTQLILQMLCAGYPGR 235
DB 171 TGHGVAPSVSLTPKLEQLEVLISRETNCNLYNIDAKPEEPHFVQEDMVCAGYVEGG 230

QY 236 RDTQCGDSGGPLVCEGGRWFOAGITSTFGCGRRNRPGVFTAVATYBAWIREQVMGSEP 295
DB 231 KDACQDSGGPLSCFVEGLWYLTGIVSWGDCAGARRPGVYTLASSYASWIOSKV--TEL 288
QY 296 GFAFTQPKTQSD---CLHQTAFIDS-ARILLRPLSHISGVSGTGTSLVLPWISPH 349
DB 289 QPRVVPTQESQPDNSNLGSHLAFSSAPAQGLLRDILFLPLGLALG---LLSEPLWSEH 343
RESULT 13
ADD47561
ID ADD47561 standard; protein; 343 AA.
XX AC ADD47561;
XX DT 29-JAN-2004 (first entry)
XX DE Human Protein Q16651, SEQ ID NO 13257.
XX KW Human; pain; neuronal tissue; gene therapy;
XX KW spinal segmental nerve injury; chronic constriction injury; CCI;
XX KW spared nerve injury; SNI; Chung.
XX OS Homo sapiens.
XX PN WO2003016475-A2.
XX PD 27-FEB-2003.
XX PF 14-AUG-2002; 2002WO-US025765.
XX PR 14-AUG-2001; 2001US-0312147P.
XX PR 01-NOV-2001; 2001US-0346382P.
XX PR 26-NOV-2001; 2001US-0333347P.
XX PA (GHEO) GEN HOSPITAL CORP.
XX PA (FARB) BAYER AG.
XX PI Woolf C, D'urso D, Befort K, Costigan M;
XX DR WPI; 2003-268312/26.
XX DR GENBANK; Q16651.
XX PT New composition comprising two or more isolated polypeptides, useful for
XX PT preparing a medicament for treating pain in an animal.
XX PS Claim 1; Page; 1017pp; English.
XX CC The invention discloses a composition comprising two or more isolated rat
XX CC or human polynucleotides or a polynucleotide which represents a fragment,
XX CC derivative or allelic variation of the nucleic acid sequence. Also
XX CC claimed are a vector comprising the novel polynucleotide, a host cell
XX CC comprising the vector, a method for identifying a nucleotide sequence
XX CC which is differentially regulated in an animal subjected to pain and a
XX CC kit to perform the method, an array, a method for identifying an agent
XX CC that increases or decreases the expression of the polynucleotide sequence
XX CC that is differentially expressed in neuronal tissue of a first animal
XX CC subjected to pain, a method for identifying a compound which regulates
XX CC the expression of a polynucleotide sequence which is differentially
XX CC expressed in an animal subjected to pain, a method for identifying a
XX CC compound that regulates the activity of one or more of the
XX CC polynucleotides, a method for producing a pharmaceutical composition, a
XX CC method for identifying a compound or small molecule that regulates the
XX CC activity in an animal of one or more of the polypeptides given in the
XX CC specification, a method for identifying a compound useful in treating
XX CC pain and a pharmaceutical composition comprising the one or more
XX CC polypeptides or their antibodies. The polynucleotide or the compound that
XX CC modulates its activity is useful for preparing a medicament for treating
XX CC pain (e.g. spinal segmental nerve injury (Chung), chronic constriction
XX CC injury (CCI) and spared nerve injury (SNI)) in an animal (e.g. gene
XX CC therapy). The sequence presented is a human protein (shown in Table 2 of
XX CC the specification) which is differentially expressed during pain. Note:

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 25, 2004, 15:37:18 ; Search time 45 Seconds
(without alignments)
763.119 Million cell updates/sec

Title: US-10-037-417-46

Perfect score: 1953

Sequence: 1 MAQKGVLPGLGAVANSDS.....TKSLVLPWLSPHSLGLMGF 357

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : PIR_78.*

1: pir1.*

2: pir2.*

3: pir3.*

4: pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	690	35.3	343	1 A57014	proctasin (EC 3.4.21.1)
2	530	27.1	855	2 JCT731	membrane-bound arg
3	514.5	26.3	275	2 A32410	tryptase (EC 3.4.21.1)
4	499	25.6	276	2 A38654	mast cell proteina
5	496	25.4	1019	1 A56318	enteropeptidase (E
6	491.5	25.2	274	2 JCA171	tryptase (EC 3.4.21.1)
7	490.5	25.1	417	1 S00845	hepsin (EC 3.4.21.1)
8	485.5	24.9	275	2 A35863	tryptase (EC 3.4.21.1)
9	484.5	24.8	275	2 B35863	tryptase (EC 3.4.21.1)
10	482.5	24.7	638	1 KOHUP	plasma kallikrein
11	481.5	24.7	274	2 A5754	tryptase (EC 3.4.21.1)
12	480	24.6	1034	1 A53663	enteropeptidase (E
13	478.5	24.5	275	2 C35863	tryptase (EC 3.4.21.1)
14	478.5	24.5	416	1 S33777	hepsin (EC 3.4.21.1)
15	475	24.3	638	1 KOMSPL	plasma kallikrein
16	474.5	24.3	270	2 S56160	mast cell tryptase
17	473.5	24.2	418	2 A37344	acrosin (EC 3.4.21.1)
18	473	24.2	421	2 S29599	acrosin (EC 3.4.21.1)
19	473	24.2	638	1 KQRTPL	plasma kallikrein
20	469.5	24.0	1035	1 A43090	enteropeptidase (E
21	469	24.0	436	2 JX0172	acrosin (EC 3.4.21.1)
22	467.5	23.9	431	2 S47538	acrosin (EC 3.4.21.1)
23	465.5	23.8	273	2 A47246	tryptase (EC 3.4.21.1)
24	464.5	23.8	269	2 B32410	mastocytoma protei
25	464.5	23.8	912	1 PLBO	plasmin (EC 3.4.21.1)
26	462	23.7	415	1 A34170	acrosin (EC 3.4.21.1)
27	460.5	23.6	437	2 S18407	acrosin (EC 3.4.21.1)
28	454	23.2	455	2 A61545	plasmin (EC 3.4.21.1)
29	452	23.1	460	2 B61545	plasmin (EC 3.4.21.1)

low-density lipopr
acrosin (EC 3.4.21.1)
plasmin (EC 3.4.21.1)
coagulation factor
plasmin (EC 3.4.21.1)
plasmin (EC 3.4.21.1)
plasmin (EC 3.4.21.1)
coagulation factor
plasmin (EC 3.4.21.1)
serine proteinase
polyprotein - Afri
pancreatic elastas
pancreatic elastas
tryptase (EC 3.4.21.1)
testicular serine
acrosin (EC 3.4.21.1)

ALIGNMENTS

RESULT 1

A57014

proctasin (EC 3.4.21.1) precursor - human

C:Species: Homo sapiens (man)

C>Date: 24-May-1996 #sequence_revision 24-May-1996 #text_change 21-Apr-2003

C:Accession: A57014; A54866

R.Yu, J.X.; Chao, L.; Chao, J.

J. Biol. Chem. 270, 13483-13489, 1995

A>Title: Molecular cloning, tissue-specific expression, and cellular localization of huma

A:Reference number: A57014; MUID:95286644; PMID:7768952

A:Accession: A57014

A>Status: translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-343 <RES>

A:Cross-references: GB:L41351; NID:9862304; PIDN:AAC41759.1; PID:9862305

A:Experimental source: prostate

A>Note: parts of this sequence were determined by protein sequencing

R.Yu, J.X.; Chao, L.; Chao, J.

J. Biol. Chem. 269, 18843-18848, 1994

A>Title: Proctasin is a novel human serine proteinase from seminal fluid. Purification, t

A:Reference number: A54866; MUID:94308140; PMID:8034638

A:Accession: A54866

A:Molecule type: protein

A:Residues: 45-64 <YUA>

C:Genetics:

A:Gene: GDB:PRSS8

A:Cross-references: GDB:676446; OMIM:600823

A:Map position: 16p11.2-16p11.2

C:Superfamily: trypsin; trypsin homology

C:Keywords: glycoprotein; hydrolase; serine proteinase; transmembrane protein

F:1-32/Domain: signal sequence #status predicted <SIG>

F:33-44/Domain: signal sequence #status predicted <MAT>

F:33-44/Domain: proctasin light chain #status predicted <CHL>

F:45-343/Domain: proctasin heavy chain #status predicted <CHH>

F:45-281/Domain: trypsin homology <TRY>

F:323-341/Domain: transmembrane #status predicted <TM1>

F:37-154/70-86/168-244/201-223/234-262/Disulfide bonds: #status predicted

F:85/134/236/Active site: His, Asp, Ser #status predicted

F:159/Binding site: carbohydrate (Asn) (covalent) #status experimental

Query Match 35.3%; Score 690; DB 1; Length 343;

Best Local Similarity 43.6%; Pred. No. 1.1e-46;

Matches 156; Conservative 49; Mismatches 129; Indels 24; Gaps 10;

QY 1 MAQKGVLPGLGAVANSDSYSLY-GLVPSGP-----ARGPYCGRPSPSARIVGSSNAQP 55

DB 1 MAQKGVLPGLGAVANSDSYSLY-GLVPSGP-----ARGPYCGRPSPSARIVGSSNAQP 53

QY 56 GTWFWQVSLHGGHICGGSLIAPSWLSAAHCFMTNGTLPAAEWSVLLGVHSDGGLD 115

DB 54 GQWFWQVSLHGGHICGGSLIAPSWLSAAHCFMTNGTLPAAEWSVLLGVHSDGGLD 110

```
QY 116 GAHTFAVAIVVPANYSQVELGADLALLRLASPASLGPAVWPVCLPRASHRFVHGTAQWA 175
Db 111 DAKVSTLKDIIIPHPSYLQEGSGDIALQLGRPIITFSYIRIPICLPANAFPNGLHCTV 170
QY 176 TQWGVQVQADPLPLPWVLQVEIRLLGATCCQLYSQPGPNLTLQILFGMLCAGYPEGR 235
Db 171 TQWGHVARSVSLITPKPLQLEVPILSRETCLNIDAKPEPFVQEDMVACAGYVGG 230
QY 236 RDTCCGDSGLPVCEBEGGRWFQAGITSGFGGRRNRPGVFTAVATYEAIRQVMGSEP 295
Db 231 KDACCQDSGLPSCPVEGLWYLTGIVSGDACARNRPGVYTLASSYASWIOSKV--TEL 288
QY 296 GFAPPTPOKQSD---CLHQTAFLDS-ARILLRLSHLSVGVSTGKSLVLPWLSPH 349
Db 289 QFRVVPVQTESQDNLCSGLHAFSAPAGQLRPLFLPLGLALG---LLSPWLSH 343

RESULT 2
JC7731
membrane-bound arginine-specific serine proteinase precursor - rat
C:Species: Rattus norvegicus (Norway rat)
C:Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 03-Feb-2003
C:Accession: JC7731; JC7775
R:Kishi, K.; Yamazaki, K.; Yasuda, I.; Yahagi, N.; Ichinose, M.; Tsuchiya, Y.; Athauda,
J. Biochem. 130, 425-430, 2001
A:Title: Characterization of a membrane-bound arginine-specific serine proteinase from rat
A:Reference number: JC7731; MUID:21421307; PMID:11530019
A:Accession: JC7731
A:Molecule type: mRNA
A:Residues: 1-855 <KIS>
A:Cross-references: DDBJ:AB049189
A:Experimental source: Strain Male, 7-week-old
R:Satomi, S.; Yanasaki, Y.; Tsuzuki, S.; Hitomi, Y.; Iwanaga, T.; Fushiki, T.
Biochem. Biophys. Res. Commun. 287, 995-1002, 2001
A:Title: A role for membrane-type serine protease (MT-SPI) in intestinal epithelial turn
A:Reference number: JC7775; PMID:11573963
A:Contents: Small intestine
A:Accession: JC7775
A:Molecule type: mRNA
A:Residues: 1-855 <SAT>
A:Cross-references: DDBJ:AB037898
C:Comment: This enzyme, an epithelial-derived, type II integral membrane serine protease
of specific proteins or peptides on the brushborder membranes. It also participates in
lial migration and/or cell loss.
C:Genetics:
A:Gene: mt-spi
A:Map position: basolateral cell surface
C:Superfamily: membrane-bound arginine-specific serine proteinase
C:Keywords: protein digestion

Query Match 27.1%; Score 530; DB 2; Length 855;
Best Local Similarity 39.9%; Pred. No. 1e-33;
Matches 112; Conservative 40; Mismatches 95; Indels 34; Gaps 8;

QY 33 RGPYPC-GRPEPS-----ARIVGSSNAQPGTPWPQVSLHH-GGSHIC 72
Db 582 KGNPECDKDKDSCDGSDEKNCDCGLRFTKQARVVGGINADEGWPQVSLHALGQGHLC 641
QY 73 GGSLLIAPSWLSAAHCFMTNGTLEPA--AEWSVLGVSQD-GPLDGHATFAVAIVVPA 129
Db 642 GASLISPDWLSAAHCFODETFKYSOHTWMTAFGLDQSKRSASGVQEHKRLIITHP 701
QY 130 NYSQVELGADLALLRLASPASLGPAVWPVCLPRASHRFVHGTAQWAGVQVQADPLPL 189
Db 702 SFNDFTFDYDIALLEKPAVSTVVRPICLPDNTHTVFPAGKALWTVGWGHTKGGTGAL 761
QY 190 PWVLQVEIRLLGATCCQLYSQPGPNLTLQILFGMLCAGYPEGRDTCQDSDGGPLV-V 248
Db 762 --ILQKGIRVINQTTCELLPQ-----QITPRMVCVGLSGVDSCQDSDGGPLSS 811
QY 249 CEEGRWFQAGITSGFGGRRNRPGVFTAVATYEAIRQVMGSEP 289
Db 812 VEKDGRIQAGVSVWGEGCAQRNKPQVYTRIPEVRDRIKEQ 852
```

RESULT 3

```
A32410
trypsinase (EC 3.4.21.59) precursor - dog
C:Species: Canis lupus familiaris (dog)
C:Date: 12-Oct-1989 #sequence_revision 12-Oct-1989 #text_change 22-Jun-1999
C:Accession: A32410
R:Vanderslice, P.; Craik, C.S.; Nadel, J.A.; Caughey, G.H.
Biochemistry 28, 4148-4155, 1989
A:Title: Molecular cloning of dog mast cell trypsinase and a related protease: structural
A:Reference number: A32410; MUID:89352460; PMID:2504277
A:Accession: A32410
A:Molecule type: mRNA
A:Residues: 1-275 <VAN>
A:Cross-references: GB:M24664; NID:g163982; PIDN:AAA30854.1; PID:g163983; GB:J02862
C:Superfamily: trypsin; trypsin homology
C:Keywords: hydrolase; serine proteinase; zymogen
F:1-21/Domain: signal sequence #status predicted <SIG>
F:22-30/Domain: activation peptide #status predicted <ACT>
F:31-275/Product: trypsinase #status predicted <MAT>
F:31-267/Domain: trypsin homology <TRY>
F:74,121,224/Active site: His, Asp, Ser #status predicted
```

```
Query Match 26.3%; Score 514.5; DB 2; Length 275;
Best Local Similarity 42.2%; Pred. No. 4.8e-33;
Matches 124; Conservative 27; Mismatches 100; Indels 43; Gaps 9;

QY 23 LYGLVPSGARGPPYCGRPEPSARIVGSSNAQPGTPWPQVSLHHGG---GHICGSLIAP 79
Db 12 LGSLLVPSFAP-----GQALQKRVGIVGGREAPCGSKWPQVSLRLKQYWRHICGSLIHP 66
QY 80 SWVLSAAHCFMTNGTLEPAAEWSVLL---GVHSQDPLDGAHTRAVAAIVVPANYSQVEL 136
Db 67 QNVLTAAHCVGN--VVCPEELRVQLREQHLYQD-----HLLPVRIVHPNYYTPEN 118
QY 137 GADLALLRLASPASLGPAVWPVCLPRASHRFVHGTAQWAGVQVQADPLPLPWVLEQV 196
Db 119 GADIALLEDEFNVSARHVQVTLPLPQTFPTGTPCWTGMDVHSGTGLPPLPPFLKQV 178
QY 197 EURLIGEATCCQLYSQPGPNLTLQILPG-----MLCAGYPEGRDTCQDSDGGPLVC 249
Db 179 KPIVENSMDVQY-----HLGLSTGCDGVIRVREDMLCAG--NSKSDSCQDSDGGPLVC 230
QY 250 EEGGRWFQAGITSGFGGRRNRPGVFTAVATYEAIRQVMGSEPGPAFTQP 303
Db 231 RVRGVWLQAGVSVWGEGCAQRNPGIYTRVAYYLDWIHQVY-----PKEP 275
```

RESULT 4

```
A38654
mast cell proteinase 6 (EC 3.4.21.-) precursor - mouse
C:Species: Mus musculus (house mouse)
C:Date: 21-Feb-1992 #sequence_revision 17-Feb-1994 #text_change 22-Jun-1999
C:Accession: A38654; B38654; D35646; I59478
R:Reynolds, D.S.; Gurley, D.S.; Austen, K.F.; Serafin, W.E.
J. Biol. Chem. 266, 3847-3853, 1991
A:Title: Cloning of the cDNA and gene of mouse mast cell protease-6. Transcription by p
A:Reference number: A38654; MUID:91139682; PMID:1995638
A:Accession: A38654
A:Molecule type: DNA
A:Residues: 1-276 <REY>
A:Cross-references: GB:M57625; NID:g200506; PIDN:AAA39987.1; PID:g200507
A:Note: The authors translated the codon CGC for residue 24 as Ala, and GAA for residue 37
s Gly, GAG for residue 148 as Gly, GAG for residue 168 as Gly, and GAA for 185 as Gly
A:Accession: B38654
A:Molecule type: mRNA
A:Residues: 1-276 <RE2>
A:Cross-references: GB:M57626; NID:g200508; PIDN:AAA39988.1; PID:g200509
R:Reynolds, D.S.; Stevens, R.L.; Lane, W.S.; Carr, M.H.; Austen, K.F.; Serafin, W.E.
Proc. Natl. Acad. Sci. U.S.A. 87, 3230-3234, 1990
A:Title: Different mouse mast cell populations express various combinations of at least
A:Reference number: A38646; MUID:90222202; PMID:2326280
```


Query Match	25.4%;	Score 496;	DB 1;	Length 1019;
Best Local Similarity	39.2%;	Pred. No. 5.6e-31;		
Matches 102;	Conservative	47;	Mismatches 85;	Indels 26;
Gaps				97
QY	38	CGR-----PEPSARIVGSSNAQGTWPMQVSLHHGGHICGSLIAPSWVLSSAAHCFMTWG	93	
Db	772	CGKLLAAQDITPKIVGSSNAKEGAPWVGIVYGRLLCGASLVSSDMLVSSAAHC-VYGR	830	
QY	94	TLPEAAEWSVLGVHSDGDLGAAHT--RAVAATVVPANYSOVELGADLALLRLASPASL	151	
Db	831	NLEP-SKWTALLGHMKS-N-LTSPQTVPLDEIVPHNRRRKDNDIAMHLEFKVNY	888	
QY	152	GPAYVPYCLPRASHRFVHGTACNATGWDV----OEADPLPLPWLOKVELRLLEATCQ	207	
Db	889	TDYIQCPLPEENQVFPFGRNCISAGTGVVYQGTAN-----ILQADYPLLSNERCQ	942	
QY	208	CLYSQPGFNLTLQILPGLCAGYPEGRRDTCQDSDGGPLVCEEGGRWFQAGITSFGFGC	267	
Db	943	---QOMPEYNIT-----ENMICAGYEEGGIDSCQDSDGGPLCMQENNRWFLAGVTSFGYKC	995	
QY	268	GRNRPGVFTAVATYEAWIR	287	
Db	996	ALPNRPGVIARVSRFTWIO	1015	

RESULT 6

RESULT 6
JC4171
trypsinase (EC 3.4.21.59) precursor - rat
N:Alternate names: mast cell trypsinase
C:Species: Rattus norvegicus (Norway rat)
C:Date: 27-Aug-1995 #sequence_revision 27-Oct-1995 #text_change 20-Jun-2000
C:Accession: JC4171
R:Idte, H.; Itoh, H.; Tomita, M.; Murakumo, Y.; Kobayashi, T.; Maruyama, H.; Osada, Y.;
J. Biochem. 118, 210-215, 1995
A:Title: cDNA sequencing and expression of rat mast cell trypsinase.
A:Reference number: JC4171; MUID:96015171; PMID:8537314
A:Accession: JC4171
A:Molecule type: mRNA
A:Residues: 1-274 <IDE>
A:Cross-references: DBJ:P38455; NID:9556555; PIDN:BAA07486.1; PID:9556556
C:Comment: this enzyme is basically specific for a connective tissue mast cell, it is up
enzyme inhibitors.
C:Superfamily: trypsin; trypsin homology
C:Keywords: glycoprotein; hydrolase; mast cell; serine proteinase; zymogen
F:1-19/Domain: signal sequence #status predicted <SIG>
F:20-29/Domain: activation peptide #status predicted <ACT>
F:30-274/Product: mast cell trypsinase #status predicted <MAT>

F:30-266/Domain: trypsin homology <TRY> F:73,120,223/Active site: His, Asp, Ser #status predicted F:131/Binding site: carbohydrate (Asn) (covalent) #status predicted		Query Match 25.2%; Score 491.5; DB 2; Length 274; Best Local Similarity 40.2%; Pred. No. 3e-31; Matches 106; Conservative 28; Mismatches 87; Indels 43; Gaps 6;	
QY	47	IVGGSNAQPGTWQVSLHHGGG---HICGGSLIAPSWLSAAHCFMNGTLEPAAEWSV	103
Db	30	IVGGRASASKPWQVSLRFKFSFWMHFCGSLIHPOWLTAAHC-----	74
QY	104	LLGVHSQDGLDGAHTR-----AAVALVVPANYSQVELGADIALLLASPASLG	152
Db	75	-VGLHHKSELFRVQREQLYVADQLLVNRVTHPHYTTVEDGADIALLELEIPNVVS	133
QY	153	PAWVPVCLPRASHRFVHGACMATGVDQVADPLPLPWQLQVEVRLLEGEATCCQCY--	210
Db	134	THIHPLSLPPASSETPPSGTSCWWTGWDIDSDPELLPPLPVKQKVPIVENSLCDKXHT	193
QY	211	-----SOPGFENLTQLPLPGLMAGYEGEGRDTCQDGGPLVCEEGGRWFQAGITSFGFG	266
Db	194	GLYTGDVPI-----VQDGLCAG--NTRSDSCQDGGPLVCKVKGKTLQAGVWSWGG	246
QY	267	CGRRNRPVFTAVATYEAWIREQV	290
Db	247	CAENRPGIYTRYVYLDWLRHV	270
RESULT 7			
S00845			
hepsin (EC 3.4.21.-) - human			
C:Species: Homo sapiens (man)			
C>Date: 31-Dec-1988 #sequence_revision 31-Dec-1988 #text_change 18-Jun-1999			
C:Accession: S00845			
R:Leytus, S.P.; Loeb, K.R.; Hagen, F.S.; Kurachi, K.; Davie, E.W.			
Biochemistry 27, 1067-1074, 1988			
A:Title: A novel trypsin-like serine protease (hepsin) with a putative transmembrane dom			
A:Reference number: S00845; MUID:88209431; PMID:2835076			
A:Accession: S00845			
A:Molecule type: mRNA			
A:Residues: 1-417 <LEY>			
A:Cross-references: EMBL:X07732; NID:g32063; PIDN:CAA30558.1; PID:g32064			
C:Genetics:			
A:Gene: GDB:HPN; TMRSS; hepsin			
A:Cross-references: GDB:133685; OMIM:142440			
A:Map position: 19q11-19q13.2			
C:Superfamily: hepsin; trypsin homology			
C:Keywords: hydrolase; liver; serine proteinase; transmembrane protein			
F:23-45/Domain: transmembrane #status predicted <TMN>			
F:163-400/Domain: trypsin homology <TRY>			
F:188-204,291-359,322-338,349-381/Diulfide bonds: #status predicted			
F:203,257,353/Active site: His, Asp, Ser #status predicted			
Query Match 25.1%; Score 490.5; DB 1; Length 417; Best Local Similarity 39.4%; Pred. No. 5.7e-31; Matches 111; Conservative 29; Mismatches 107; Indels 35; Gaps 9;			
QY	38	CGREE-PSARIVGGSNAQPGTWQVSLHHGGCHICGSLIAPSWLSAAHCFMNTGLE	96
Db	153	CGREKLPVDRIVGGRDTSGLGRWPQVSLRYDGAHLCCGSLLSGDWVLTAAHCFERNRV-	211
QY	97	PAAEWSVLGVHSDGDLGDAH--TRAVAAIVVPANYSQVELGADIALLLASLP	148
Db	212	-LSRNRVPAGVAQASP-----HGLQLGVQAVVYHGGVLPFRDPNSENSNDIALVHLSPP	266
QY	149	ASLGFAPVPCVCLPRASHRFVHGACMATGVDV---QADPLPLPWQLQVEVRLLEGEA	204
Db	267	LPLTEYIQVCLPRAQALVDGKICTVTGWTGNTQYVQQAQ-----VLQEARVPIISND	320
QY	205	TCQCLYSQGFENLTQLPLPGLMAGYEGEGRDTCQDGGPLVCEEG---GRWFQAGI	260
Db	321	VCN-----GADFYGNQIKPMFCAGYEGGIDACQDGGSGFFVCEDSISRTPRRLCGI	374

QY	261	TSFGCGGRRNRPGVFTAVATYEAWIREQVVMGSEPPGPAFPPTQ	302
Db	375	VSWGTCALAQKPGVYKVSDFREWIFQAIKTHSEASGMVITQ	416
RESULT 8			
A35863			
trypsinase (EC 3.4.21.59) I precursor - human			
C:Species: Homo sapiens (man)			
C>Date: 12-Oct-1990 #sequence_revision 12-Oct-1990 #text_change 21-Jul-2000			
C:Accession: A35863; D35863; A60939; A39326			
R:Vanderslice, P.; Ballinger, S.M.; Tam, B.K.; Goldstein, S.M.; Craik, C.S.; Caughey, G.			
Proc. Natl. Acad. Sci. U.S.A. 87, 3811-3815, 1990			
A:Title: Human mast cell tryptase: multiple cDNAs and genes reveal a multigene serine pr			
A:Reference number: A35863; MUID:90251647; PMID:2187193			
A:Accession: A35863			
A:Molecule type: DNA			
A:Residues: 1-275 <VAN>			
A:Cross-references: GB:M33494; NID:g3927804; PIDN:AAC83172.1; PID:g339977			
A:Accession: D35863			
A:Molecule type: mRNA			
A:Residues: 1-275 <VA2>			
A:Cross-references: GB:M33491			
R:Butterfield, J.H.; Weiler, D.A.; Hunt, L.W.; Wynn, S.R.; Roche, P.C.			
J. Leukoc. Biol. 47, 409-419, 1990			
A:Title: Purification of tryptase from a human mast cell line.			
A:Reference number: A60939; MUID:90244210; PMID:2110591			
A:Accession: A60939			
A:Molecule type: protein			
A:Residues: 31-38, 'P',40-41, 'X',43, 'T',45-48, 'X',50 <BUT>			
A:Experimental source: mast cell			
A:Note: 44-Gly was also found			
R:Cronlish, J.A.; Seidah, N.G.; Marcinkiewicz, M.; Hamelin, J.; Johnson, D.A.; Chretien,			
J. Biol. Chem. 262, 1363-1373, 1987			
A:Title: Human pituitary tryptase: molecular forms, NH-2-terminal sequence, immunocytoch			
A:Reference number: A39326; MUID:87109258; PMID:3543004			
A:Accession: A39326			
A:Molecule type: protein			
A:Residues: 31-38 <CRO>			
A:Experimental source: pituitary			
C:Genetics:			
A:introns: 21/1; 78/2; 177/1; 221/3			
C:Superfamily: trypsin; trypsin homology			
C:Keywords: hydrolase; serine proteinase; zymogen			
F:1-21/Domain: signal sequence #status predicted <SIG>			
F:22-30/Domain: activation peptide #status predicted <ACT>			
F:31-275/Product: trypsin I #status experimental <MAT>			
F:31-267/Domain: trypsin homology <TRY>			
F:74,121,224/Active site: His, Asp, Ser #status predicted			
Query Match 24.9%; Score 485.5; DB 2; Length 275; Best Local Similarity 41.7%; Pred. No. 9e-31; Matches 111; Conservative 32; Mismatches 106; Indels 17; Gaps 7;			
QY	32	ARGPPVCGPEPARIVGGSNAQPGTWQVSLH-HG--GGHICGSLIAPSWLSAAHC	88
Db	16	AYAAPAGQALQVIGVGGQEAPEPSKPMQVSLRVHGMHFCGSLIHPOWLTAAHC	75
QY	89	FMINGTLEPAEWSVLL---GVHSDGDLGDAHTRAVAAIVVPANYSQVELGADIALLL	145
Db	76	--VGPDVKDLAALRVQLREQLHYQQQL-----PVSRIIVHPQFVTAQIGADIALLE	127
QY	146	ASPAISGPAVWPVCLPRASHRFVHGACMATGVDVQVADPLPLPWQLQVEVRLLEGEAT	205
Db	128	EEPNVSSVHTVTLPPASSETPPGPECVWGVDVNDERLPPFPFLKQKVPIMENHI	187
QY	206	CQCLYSQGFENLTQLL-PMGLCAGYEGERRTCCGDSGGPLVCEEGGRWFQAGITSFG	264
Db	188	CDARYHLGAYTGDVIRIVDDMLCAG--NTRRSDSCQDGGPLVCKVKGKTLQAGVWSW	245
QY	265	FGCGRRNRPGVFTAVATYEAWIREQV	290

Db 246 EGCAQNPREGIYTRVYLDWIHHYV 271

RESULT 9

B35863
N:Alternate names: trypsin - human
C:Species: Homo sapiens (man)
C:Date: 03-Feb-1994 #sequence_revision 03-Feb-1994 #text_change 21-Jul-2000
C:Accession: B35863; A37193; I59473
R:Vanderslice, P.; Ballinger, S.M.; Tam, E.K.; Goldstein, S.M.; Craik, C.S.; Caughey, G.
Proc. Natl. Acad. Sci. U.S.A. 87, 3811-3815, 1990
A:Title: Human mast cell trypsinase: multiple cDNAs and genes reveal a multigene serine pr
A:Reference number: B35863; MUID:90251647; PMID:2187193
A:Accession: B35863
A:Molecule type: mRNA; DNA
A:Residues: 1-275 <VAN>
A:Cross-references: GB:M33492; NID:g339982; PIDN:AAA36779.1; PID:g339983
A:Note: residues 2-275 are derived from mRNA; residue one was inferred from the genomic
R:Miller, J.S.; Moxley, G.; Schwartz, L.B.
J. Clin. Invest. 86, 864-870, 1990
A:Title: Cloning and characterization of a second complementary DNA for human trypsinase.
A:Reference number: A37193; MUID:90369005; PMID:2203827
A:Accession: A37193
A:Molecule type: mRNA
A:Residues: 1-275 <MIL>
A:Cross-references: GB:M37488; NID:g179583; PIDN:AAA51843.1; PID:g179584
R:Blom, T.; Hellman, L.
Scand. J. Immunol. 37, 203-208, 1993
A:Title: Characterization of a trypsinase mRNA expressed in the human basophil cell line K
A:Reference number: I59473; MUID:93166209; PMID:8434231
A:Accession: I59473
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-275 <RES>
A:Cross-references: GB:S5551; NID:g265666; PIDN:AAD13876.1; PID:g4261576
A:Experimental source: basophil cell line K0812
C:Genetics:
A:Gene: GDB:TPS1
A:Cross-references: GDB:125890; OMIM:191080
A:Map position: lepton-16qter
C:Superfamily: trypsin; trypsin homology
C:Keywords: hydrolyase; serine proteinase; zymogen
F:1-21/Domain: signal sequence #status predicted <SIG>
F:22-30/Domain: activation peptide #status predicted <ACT>
F:31-275/Domain: trypsin homology <TRY>
F:31-267/Domain: trypsin homology <TRY>
F:74,121,224/Active site: His, Asp, Ser #status predicted

Query Match 24.8%; Score 484.5; DB 2; Length 275;
Best Local Similarity 41.7%; Pred. No. 1.1e-30;
Matches 111; Conservative 31; Mismatches 107; Indels 17; Gaps 7;
Qy 32 ARGPPYCORPEPSARIYVGGNAQGTWPQVSLH-HG--GGHICGSLIAPSWLVAHAC 88
Db 16 AYAAPAPQALQRYGVIGQGPAPRSKMPQVSLRVHGPYHMFHFCGSLIHPQWVLTAAHC 75
Qy 89 FWTNGTLEPAEWSVLL--GVHSODGLDCAHTRAAVAVVPANYSQVELGADLALLRL 145
Db 76 --VGPDKDLAALRVQLEQHLVQDQL-----PVSRIVHQPQFYTAQIADLALLEL 127
Qy 146 ASPASIGPAPVFPVCLPRASHRFVHGTCATWGVDQVQADPLPLPWLQVYELRLLEGAT 205
Db 128 EEPVKVSHVHTVTLPPASETFPPGMPQWVTGWDVNDERLPPPLKQVKVPMENHI 187
Qy 206 CQCLYSQGPFLNLTQIL-PCMLCAGYEGRRDTCQSGGGLVCEEGRRWFQAGITSPG 264
Db 188 CDAKYHLGATGDDVRIYRDDMLCAG--NTRRDSQCGSGGLVCKVNGTQLQAGVSWG 245
Qy 265 FCGGRNRPVGFATVATYEAWIREQV 290
Db 246 EGCAQNPREGIYTRVYLDWIHHYV 271

RESULT 10

KOHUP
N:Alternate names: kininogenin; plasma prekallikrein
C:Species: Homo sapiens (man)
C:Date: 13-Aug-1986 #sequence_revision 13-Aug-1986 #text_change 18-Jun-1999
C:Accession: A00921; A37939
R:Chung, D.W.; Fujikawa, K.; McMullen, B.A.; Davie, E.W.
Biochemistry 25, 2410-2417, 1986
A:Title: Human plasma prekallikrein, a zymogen to a serine protease that contains four te
A:Reference number: A00921; MUID:86243359; PMID:3521732
A:Accession: A00921
A:Molecule type: mRNA
A:Residues: 1-638 <CHU>
A:Cross-references: GB:M13143; NID:g190262; PIDN:AAA60153.1; PID:g190263
R:McMullen, B.A.; Fujikawa, K.; Davie, E.W.
Biochemistry 30, 2050-2056, 1991
A:Title: Location of the disulfide bonds in human plasma prekallikrein: the presence of f
A:Reference number: A37939; MUID:91152016; PMID:1998666
A:Accession: A37939
A:Molecule type: protein
A:Residues: 20-27;40-46,'X',48,'H',50,'X',52-70,'H',75-76,'X',78-80;103-113;131-140;141-1
;260-283,'X',285;287-291,'X',293-295;314-317,'X',319-320;321-324,'X',329-333;334-339,'X',
525;538-551;562,'X',564-567;573,'X',575-576;578-583,'X',585;592-604 <WCM>
C:Comment: This protein, synthesized in the liver, circulates as a noncovalent complex w
C:Comment: The zymogen is activated by factor XIIa, which cleaves the molecule into a lig
are linked by one or more disulfide bonds.
C:Comment: The enzyme cleaves Lys-Arg and Arg-Ser bonds. It activates, in a reciprocal re
inogen and may also play a role in the renin-angiotensin system by converting prorenin in
C:Genetics:
A:Gene: GDB:KLK3
A:Cross-references: GDB:127575; OMIM:229000
A:Map position: 4q35-4q35
C:Superfamily: coagulation factor XI; trypsin homology
C:Keywords: blood coagulation; duplication; fibrinolysis; glycoprotein; hydrolase; inflan
F:1-19/Domain: signal sequence #status predicted <SIG>
F:20-638/Product: plasma kallikrein #status predicted <MAT>
F:20-390/Domain: plasma kallikrein heavy chain #status predicted <HCH>
F:110-199/Domain: apple repeat <AP1>
F:110-199/Domain: apple repeat <AP2>
F:200-289/Domain: apple repeat <AP3>
F:291-380/Domain: apple repeat <AP4>
F:391-621/Domain: trypsin homology <TRY>
F:21-104,47-77,51-57,111-194,137-166,141-147,201-284,227-256,231-237,292-375,322-328,383-
F:127,308,396,453,494/Binding site: carbohydrate (Asn) (covalent) #status experimental
F:318-347,340-345/Disulfide bonds: #status predicted
F:390-391/Cleavage site: Arg-Ile (coagulation factor XIIa) #status predicted
F:434,483,578/Active site: His, Asp, Ser #status predicted

Query Match 24.7%; Score 482.5; DB 1; Length 638;
Best Local Similarity 40.2%; Pred. No. 3.9e-30;
Matches 104; Conservative 33; Mismatches 97; Indels 25; Gaps 6;
Qy 44 SARIYVGSNAQGTWPQVSLH---HGGHICGSLIAPSWLVAHACFWNGTLEPAAE 100
Db 388 STRIVGTNSWGEWPPQVSLQVKTQAQRLCGSLIGHQWLTAAHCF---DGLPLQDV 444
Qy 101 WSVLLGV-----HSQDGLDCAHTRAAVAVVPANYSQVELGADLALLRLASPSLGPV 155
Db 445 WRIYSGILNLSDTKOTPPS-----QIKELIHHONYKVSQGNHDIALLKQAPLNYTEFQ 499
Qy 156 WPVCLPRASHRFVHGTCATWGVDQVQADPLPLPWLQVYELRLLEGATCQCLYSQGP 215
Db 500 KPICLPSKGDSTIYINCWVTGWFSGSKGKGIQN--ILQKVNIPLVNTEECQKRYQD--- 554
Qy 216 FNLTQLPLPMLCAGYEGRRDTCQSGGGLVCEEGRRWFQAGITSGFGGRNRPV 275
Db 555 ----YKLTQRMVCAKYEGGKQACKGDSGGGLVCKHGMRLVLTGTSWGEGCARREQPGV 610
Qy 276 FTAVATYEAWIREQVMSGE 294

Db 611 YTKVAEYMDWILEKXQSSD 529

RESULT 11

A45754
trypsinase (EC 3.4.21.59) alpha precursor - human
C:Species: Homo sapiens (man)
C>Date: 03-Jun-1993 #sequence_revision 03-Jun-1993 #text_change 08-Sep-1997
C:Accession: A45754; B37193
R:Miller, J.S.; Westin, E.H.; Schwartz, L.B.
J. Clin. Invest. 84, 1188-1195, 1989
A:Title: Cloning and characterization of complementary DNA for human trypsinase.
A:Reference number: A45754; MUID:90009311; PMID:2677049
A:Accession: A45754
A:Molecule type: mRNA
A:Residues: 1-274 <MIL>
A:Cross-references: GB:M30038
R:Miller, J.S.; Moxley, G.; Schwartz, L.B.
J. Clin. Invest. 86, 864-870, 1990
A:Title: Cloning and characterization of a second complementary DNA for human trypsinase.
A:Reference number: A37193; MUID:90369005; PMID:2203827
A:Accession: B37193
A:Molecule type: mRNA
A:Residues: 1-274 <M12>
A:Cross-references: GB:M30038
A:Note: the sequence from Fig. 4 is inconsistent with that from Fig. 2 in having 205-Pro
C:Superfamily: trypsin; trypsin homology
C:Keywords: hydrolase; serine proteinase; zymogen
F:1-21/Domain: signal sequence #status predicted <SIG>
F:22-30/Domain: activation peptide #status predicted <ACT>
F:31-274/Product: trypsinase I #status predicted <MAT>
F:31-266/Domain: trypsin homology <TRY>
F:74,120,223/Active site: His, Asp, Ser #status predicted
Query Match 24.7%; Score 481.5; DB 2; Length 274;
Best Local Similarity 40.8%; Pred. No. 1.8e-30;
Matches 108; Conservative 31; Mismatches 97; Indels 29; Gaps 7;
QY 41 PEP-----SARIVGGNAQPTWQVSLH--HGCGHICGGSLLIAPSWLVAHCFMIN 92
Db 20 PAPVQALQAGIVGGQFAPSKFWQVSLRVDRYWHFCCGSLHPQWVLTIANHC----- 75
QY 93 GTEPEAAEWSVLGVHSQDGLDGAH-----TRAVAAIIVVPANYSQVELGADLALLRLA 146
Db 76 --LGPDKOLATRVNS-----GTHLYYQDQLLPVSRIMVHPQFVLIQTGADIALLELE 127
QY 147 SPASLGPVWPVCLPRASHRFVHGTCATWGQVQADPLPLPWVLOEVELLGEATC 206
Db 128 EPVNISSRVHTVMLPPASSETFPFGMPCWVTGWGDVDNDELPFPFLPKQVKVPMENHIC 187
QY 207 QCLYSQPGPNLTQLIL-PCMLCAGYEPGRRDTCQDGGPLVCEGGRWFQAGITSFGF 265
Db 188 DAKHLGAYTGDDVRIIRDMLCAG--NSQDSCKSGSGPLVCKYNGTWLQAGVSWDE 245
QY 266 GCGRRNRPVFTAVATYEAIRQV 290
Db 246 GCAQNPNGIVTRYVYLDLMIHV 270
RESULT 12
A53663
enteropeptidase (EC 3.4.21.9) precursor [validated] - pig
N:Alternate names: enterokinase
C:Species: Sus scrofa domestica (domestic pig)
C>Date: 07-Oct-1994 #sequence_revision 09-Aug-1996 #text_change 28-Apr-2003
C:Accession: A53663
R:Matsushima, M.; Ichinose, M.; Yahagi, N.; Kakei, N.; Tsukada, S.; Miki, K.; Kurokawa, J. Biol. Chem. 269, 19976-19982, 1994
A:Title: Structural characterization of porcine enteropeptidase.
A:Reference number: A53663; MUID:94327548; PMID:8051081
A:Accession: A53663
A:Molecule type: mRNA
A:Residues: 1-1034 <MAT>

A:Cross-references: GB:D30799; NID:g505122; PIDN:RAA06459.1; PID:g505123
A:Note: parts of this sequence, including the amino ends of three chains isolated from t
C:Comment: The mechanism of association with the membrane of the intestinal brush border
ated below) or with amino-terminal myristoylation of the heavy chain.
C:Complex: Mature enteropeptidase is variously reported to contain two (heavy and light)
ed by a disulfide bond. Possibly, conversion from membrane-bound to soluble forms involv
C:Function:
A:Description: cleaves activation peptide from trypsinogen to produce active trypsin
A:Pathway: intestinal digestive hydrolase cascade
C:Superfamily: enteropeptidase; Ctr/Cls repeat homology; LDL receptor ligand-binding ref
C:Keywords: glycoprotein; hydrolase; serine proteinase; transmembrane protein; zymogen
F:22-38/Domain: transmembrane #status predicted <TM>
F:52-117/Product: enteropeptidase mini chain #status predicted <MCH>
F:118-799/Product: enteropeptidase heavy chain #status predicted <HCH>
F:199-236/Domain: LDL receptor ligand-binding repeat homology <LDL1>
F:357-519/Domain: MAM homology <MAM>
F:541-646/Domain: Ctr/Cls repeat homology <CTR>
F:658-692/Domain: LDL receptor ligand-binding repeat homology <LDL2>
F:693-798/Domain: scavenger receptor cysteine-rich domain homology #status atypical <SRC
F:800-1034/Product: enteropeptidase light chain #status predicted <LCH>
F:800-1029/Domain: trypsin homology <TRY>
F:116,147,170,194,283,343,350,403,455,485,518,549,645,697,701,721,740,761,804,863,902,96
F:787-911,825-841,925-992,956-971,982-1010/Disulfide bonds: #status predicted
F:840,891,986/Active site: His, Asp, Ser #status predicted

Query Match 24.6%; Score 480; DB 1; Length 1034;

Best Local Similarity 38.5%; Pred. No. 1e-29;
Matches 99; Conservative 46; Mismatches 92; Indels 20; Gaps 9;

QY 38 CGR-----PEPSARIVGGNAQPTWQVSLHGGGHHICGGSLLIAPSWLVAHCFMIN 93
Db 787 CGKQVQAEVSPKIVGGNDSREGAWPVVVALYNGQLICGASLVSRDLVLSAHC-VYGR 845
QY 94 TLEPEAAEWSVLGVHSQDGLDGAH--TRAVAAIIVVPANYSQVELGADLALLRLASPASL 151
Db 846 NLEP-SKWKAILGLHWTN-LTSPQIVTRLIDEIVNPHYNNRRKDSITAMHLEFKVNY 903
QY 152 GPVAPVWPVCLPRASHRFVHGTCATWGQDV-DEADPLPLPWVLOEVELLGEATCQCCLY 210
Db 904 TDVIOPICLPEENQVPPGPRICSIAGWGKVIYQGSFAD---ILOEADVELLSNEKQ--- 957
QY 211 SQPGPNLTQLIL-PCMLCAGYEPGRRDTCQDGGPLVCEGGRWFQAGITSFGFCGGR 270
Db 958 QQMEYNIT----ENMWCAGYEGGIDSCQDGGPLVCEGGRWFQAGITSFGFCGGR 1013
QY 271 NRPVFTAVATYEAIR 287
Db 1014 NRPVFTAVATYEAIR 1030

RESULT 13

C35863
trypsinase (EC 3.4.21.59) III precursor - human
C:Species: Homo sapiens (man)
C>Date: 03-Feb-1994 #sequence_revision 03-Feb-1994 #text_change 07-Mar-2003
C:Accession: C35863; E35863; A38893
R:Vanderslice, P.; Ballinger, S.M.; Tam, E.K.; Goldstein, S.M.; Craik, C.S.; Caughey, G.
proc. Natl. Acad. Sci. U.S.A. 87, 3811-3815, 1990
A:Title: Human mast cell trypsinase: multiple cDNAs and genes reveal a multigene serine pr
A:Reference number: A35863; MUID:90251647; PMID:2187193
A:Accession: C35863
A:Molecule type: mRNA
A:Residues: 9-275 <VAN>
A:Accession: E35863
A:Molecule type: DNA
A:Residues: 1-9 <VA2>
A:Cross-references: GB:M33494; NID:G3927804; PIDN:AAAC83172.1; PID:G339977
A:Note: the first nine residues of this sequence are inferred from genomic DNA of trypt
submitted to GenBank, April 1990
A:Reference number: A38893
A:Accession: A38893
A:Molecule type: mRNA

A;Residues: 9-131,'K',132-275 <VA3>
A;Cross-references: GB:M33493; NID:G333984; PIDN:AAA36780.1; PID:G333985
C;Superfamily: trypsin; trypsin homology
C;Keywords: hydrolase; serine proteinase; zymogen
F;1-21/Domain: signal sequence #status predicted <SIG>
F;22-30/Domain: activation peptide #status predicted <ACT>
F;31-275/Product: trypsin I #status predicted <MAT>
F;31-267/Domain: trypsin homology <TRY>
F;74,121,224/Active site: His, Asp, Ser #status predicted

Query Match 24.5%; Score 478.5; DB 2; Length 275;
Best Local Similarity 41.0%; Pred. No. 3.2e-30;
Matches 109; Conservative 32; Mismatches 108; Indels 17; Gaps 6;

QY 32 ARGPPYCGRPESARIVGSGNAQPGTWPMQVSLH---HGGHICGSLIAPSWLSAAHC 88
DB 16 AYAAPAGQALQGVIGGCEAPRKPMPQVSLRVDRVWMEFCGSLIHPQWVITAAHC 75
QY 89 FMTNGTLEPAEAWSVLL---GVESQDGLDGAHTRAVAAIVVPANYSQVELGADLALRL 145
DB 76 --VGPDVKDLAALRVQLREOHLVYQDQLL-----PVSRIIVHPQFYTAQIGADIALLEL 127
QY 146 ASPASLGPAWVPCVLPRAHFRVHGTACWATGWDVQVQVADPLPLVWLOEVELLIGAT 205
DB 128 EEPVNVSSHVHTVTLPPASETFPPGMPVWTGWDVNDERLPPFPFLKQVVPIMENHI 187
QY 206 CQCLYSQPGFNLTLQIL-PMGLCAGYPRGRDTCQDGGPLVCBEGGRWFOAGITSPG 264
DB 188 CDAKHGLGAYTGDDVIRVDRDMLCAG--NTRDSCQDGGPLVCCKVGTWLOAGVWSWG 245
QY 265 FGCGRRNRPGVFTAVATYEAWIREQV 290
DB 246 EGCAQENRPGIYTRVYLDWIHYV 271

RESULT 14

S33777
hepsin (EC 3.4.21.-) - rat
C;Species: Rattus norvegicus (Norway rat)
C;Date: 06-Jan-1995 #sequence_revision 06-Jan-1995 #text_change 18-Jun-1999
C;Accession: S33777; S32013
R;Farley, D.; Raymond, F.; Nick, H.
Biochim. Biophys. Acta 1173, 350-352, 1993
A;Title: Cloning and sequence analysis of rat hepsin, a cell surface serine proteinase.
A;Reference number: S33777; MUID:93305733; PMID:8318546
A;Accession: S33777
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-416 <FAR>
A;Cross-references: EMBL:X70900; NID:G57928; PIDN:CAA50256.1; PID:G57929
C;Superfamily: hepsin; trypsin homology
C;Keywords: hydrolase; liver; serine proteinase; transmembrane protein
F;22-44/Domain: transmembrane #status predicted <TM>
F;162-399/Domain: trypsin homology <TRY>
F;187-203,320-358,321-337,348-380/Disulfide bonds: #status predicted
F;202,256,352/Active site: His, Asp, Ser #status predicted

Query Match 24.5%; Score 478.5; DB 1; Length 416;
Best Local Similarity 38.6%; Pred. No. 5e-30;
Matches 107; Conservative 30; Mismatches 117; Indels 23; Gaps 7;

QY 38 CGRPE-PSARIVGSGNAQPGTWPMQVSLHGGHICGSLIAPSWLSAAHCFMTNGTLE 96
DB 152 CGRKLVPDRIVGQDSSLGFWPMQVSLRYDGTGLCGSLLSGDWLTAHCFPERNRV- 210
QY 97 PAEWSVLLGVHSGQDGLDGAHTRAVAAIVVPANYSQV-----ELGADIALLASPAS 150
DB 211 -LSRWRFVAGAVARTSP--HAVQLGQAVIVHGGYLPDRPTIDENSNDIALVHLSLSP 267
QY 151 LGPAWVPCVLPRAHFRVHGTACWATGWDVQVQVADPLPLVWLOEVELLIGATCCQCLY 210
DB 268 LTVETQVPCVLPAAQALVDGKVTCTVGTGNTQFYGQQAV--VLQEARVPIISNEVCN--- 322

QY 211 SQPGFNLTLQILPMLCAGYPRGRDTCQDGGSLVCEE-----GGRWFOAGITSPFG 266
DB 323 ---SPDFYGNQIKPMKFCAGYPEGGIDACQDGGHFCVEDRISGTSRWRLCGLVSWG 379
QY 267 CGRRNRPGVFTAVATYEAWIREQVMSGSEPGPAFTQP 303
DB 380 CALARKPGVYTKVIDREWIFQAIKTHSEATGMVTP 416

RESULT 15
KQMSPL
plasma kallikrein (EC 3.4.21.34) precursor - mouse
C;Species: Mus musculus (house mouse)
C;Date: 30-Sep-1992 #sequence_revision 30-Sep-1992 #text_change 18-Jun-1999
C;Accession: A36557
R;Seldah, N.G.; Sawyer, N.; Hamelin, J.; Mion, P.; Beaubien, G.; Brachpapa, L.; Rochemont
DNA Cell Biol. 9, 737-748, 1990
A;Title: Mouse plasma kallikrein: cDNA structure, enzyme characterization, and comparison
A;Reference number: A36557; MUID:91090844; PMID:2264928
A;Accession: A36557
A;Molecule type: mRNA
A;Residues: 1-638 <SEI>
A;Cross-references: GB:M58588; NID:G200358; PIDN:AAA63393.1; PID:G200359
A;Note: part of this sequence, including the amino ends of both the heavy and light chain
C;Comment: This protein, synthesized in the liver, circulates as a noncovalent complex w
C;Comment: The zymogen is activated by factor XIIa, which cleaves the molecule into a li
are linked by one or more disulfide bonds.
C;Superfamily: coagulation factor XI; trypsin homology
C;Keywords: blood coagulation; duplication; fibrinolysis; glycoprotein; hydrolase; infla
F;1-19/Domain: signal sequence #status predicted <SIG>
F;120-390/Product: plasma kallikrein heavy chain #status experimental <HCH>
F;20-109/Domain: apple repeat <AP1>
F;110-199/Domain: apple repeat <AP2>
F;200-289/Domain: apple repeat <AP3>
F;291-380/Domain: apple repeat <AP4>
F;391-638/Product: plasma kallikrein light chain #status experimental <LCH>
F;391-621/Domain: trypsin homology <TRY>
F;21-104,47-77,51-57,111-194,137-166,141-147,201-284,227-256,231-237,292-375,318-347,322-
F;127,215,308,396,494/Binding site: carbonyl (Asn) (covalent) #status predicted
F;434,483,578/Active site: His, Asp, Ser #status predicted

Query Match 24.3%; Score 475; DB 1; Length 638;
Best Local Similarity 39.0%; Pred. No. 1.5e-29;
Matches 110; Conservative 37; Mismatches 111; Indels 24; Gaps 8;

QY 18 SDSYS--LYGLVPSGPARGPPYCGRPESARIVGSGNAQPGTWPMQVSLH--HGGHIC 72
DB 367 SSGYSRLCKLVDS-----PDC-TTKINARIVGGTNASLGEMPMQVSLQVLSQTHLC 419
QY 73 GGSIIAPSWLSAAHCFMTNGTLEPAEWSVLLGVHSGQDGLDGAHTRAVAAIVVPANYS 132
DB 420 GGSIIIGRWLTAACHF--DGIPYPDV-WRIYGGILSLSEITKETPSSRIKELIIHQEYK 476
QY 133 QVELGADLALLRLASPASLGPAWVPCVLPRAHFRVHGTACWATGWDVQVQVADPLPLPW 192
DB 477 VSEGNVDIALIKLQTPINTTEFOKPICLFSKADTNIYNCWTGHWGTYKEQGETQN--I 534
QY 193 LQVVELRLIGEATCCQLYSQPGFNLTLQILPMLCAGYPRGRDTCQDGGSLVCEE 252
DB 535 LQKATPLVPNECQKYRD-----YVINKQMICAGYKEGTDACKGDSGGPLVCKHS 587
QY 253 GRWFOAGITSPFGCGRRNRPGVFTAVATYEAWIREQVMSGE 294
DB 588 GRWLQVIGITSWGEGCGKQDQPGVYTKVSEYMDWILEKTQSSD 629

Search completed: February 25, 2004, 15:50:30

Job time : 46 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 25, 2004, 14:37:37 ; Search time 49 Seconds
(without alignments)
379.368 Million cell updates/sec

Title: US-10-037-417-46

Perfect score: 1953

Sequence: 1 MAQKGVLPQGLGAVANSDS.....TKSLVLPWLSPHSLGLGWGF 357

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_42.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	690	35.3	343	1 PSS8_HUMAN	Q16651 homo sapien
2	631	32.3	342	1 PSS8_MOUSE	Q9esd1 mus musculu
3	620	31.7	342	1 PSS8_RAT	Q9es87 rattus norv
4	582	29.8	321	1 TRYX_HUMAN	Q9nr12 homo sapien
5	563	28.8	290	1 PR27_HUMAN	Q9bqr3 homo sapien
6	547	28.0	811	1 TMS6_HUMAN	Q8iu80 homo sapien
7	544	27.9	317	1 BSS4_HUMAN	Q8gz24 homo sapien
8	540	27.6	855	1 ST14_HUMAN	Q9y5y6 homo sapien
9	534.5	27.4	314	1 TEST_HUMAN	Q9y6m0 homo sapien
10	533	27.3	811	1 TMS6_MOUSE	Q9db10 mus musculu
11	529	27.1	324	1 TEST_MOUSE	Q9jhj7 mus musculu
12	528.5	27.1	311	1 TRYX_MOUSE	Q9ql17 mus musculu
13	523	26.8	855	1 ST14_MOUSE	P56677 mus musculu
14	516	26.4	306	1 BSS4_MOUSE	Q9er10 mus musculu
15	514.5	26.3	275	1 TRYX_CANFA	P15944 canis fami
16	508.5	26.0	273	1 TRYX_SHEEP	Q9xsm2 ovis aries
17	499	25.6	275	1 TRYX_PIG	Q9n2d1 sus scrofa
18	499	25.6	276	1 MCT6_MOUSE	P21845 mus musculu
19	496	25.4	1019	1 ENTK_HUMAN	P98073 homo sapien
20	494.5	25.3	454	1 TMS3_HUMAN	P57727 homo sapien
21	494	25.3	453	1 TMS3_MOUSE	Q8k1t0 mus musculu
22	491.5	25.2	274	1 MCT6_RAT	P50343 rattus norv
23	490.5	25.1	417	1 HEPS_HUMAN	P05981 homo sapien
24	490.5	25.1	436	1 HEPS_MOUSE	Q35453 mus musculu
25	485.5	24.9	275	1 TRB1_HUMAN	Q15661 homo sapien
26	484.5	24.8	275	1 TRB2_HUMAN	P20231 homo sapien
27	482.5	24.7	638	1 ENTK_HUMAN	P01952 homo sapien
28	480	24.6	1034	1 ENTK_PIG	P98074 sus scrofa
29	479	24.5	275	1 TRYX_HUMAN	P15157 homo sapien
30	478.5	24.5	416	1 HEPS_RAT	Q95511 rattus norv
31	477.5	24.4	457	1 TMS5_HUMAN	Q9h3s3 homo sapien
32	475	24.3	638	1 KAL_MOUSE	P36262 mus musculu
33	474.5	24.3	270	1 TRYX_MERON	P50342 meriones un

RESULT 1
PSS8_HUMAN ID PSS8_HUMAN STANDARD; PRT; 343 AA.
AC Q16651; Q9UCA3;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Prostatin precursor (EC 3.4.21.-).
GN PSS8.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC TISSUE=Prostate;
RX MEDLINE=95286644; PubMed=7768952;
RA Yu J.X.; Chao L.; Chao J.;
RT "Molecular cloning, tissue-specific expression, and cellular
localization of human prostatin mRNA";
RL J. Biol. Chem. 270:13483-13489(1995).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Placenta;
RX MEDLINE=2238257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
Brownstein M.J., Udén T.B., Toshiyuki S., Carninci P., Prange C.,
Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullah S.J.,
Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
Fahey J., Helton E., Kettaman M., Madan A., Rodriguez S., Sanchez A.,
Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length
human and mouse cDNA sequences";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [3]
RP SEQUENCE OF 45-64.
RC TISSUE=Semen;
RX MEDLINE=94308140; PubMed=8034638;
RA Yu J.X.; Chao L.; Chao J.;
RT "Prostatin is a novel human serine proteinase from seminal fluid.
Purification, tissue distribution, and localization in prostate
gland";
RL J. Biol. Chem. 269:18843-18848(1994).
CC -!- FUNCTION: Possesses a trypsin-like cleavage specificity.
CC -!- SUBUNIT: Heterodimer of two chains, light and heavy, held by a

ALIGNMENTS

34	473	24.2	638	1	KAL_RAT	P14272 rattus norv
35	472	24.2	1042	1	COR1_HUMAN	Q9Y5G5 homo sapien
36	469.5	24.0	1035	1	ENTK_BOVIN	P98072 bos taurus
37	469	24.0	436	1	ACRO_MOUSE	P23578 mus musculu
38	467.5	23.9	431	1	ACRO_RABIT	P48038 oryctolagus
39	465.5	23.8	273	1	MCT7_MOUSE	Q02844 mus musculu
40	464.5	23.8	269	1	TRYX_CANFA	P19236 canis fami
41	464.5	23.8	812	1	PLMN_BOVIN	P06868 bos taurus
42	462	23.7	415	1	ACRO_PIG	P08001 sus scrofa
43	460.5	23.6	437	1	ACRO_RAT	P29293 rattus norv
44	459	23.5	273	1	MCT7_RAT	P27435 rattus norv
45	458.5	23.5	455	1	TMS5_MOUSE	Q9er04 mus musculu


```

FT PROPEP 323 342 BY SIMILARITY.
FT TRANSMEM 320 340 POTENTIAL.
FT DOMAIN 45 286 SERINE PROTEASE.
FT DISULFID 37 154 INTERCHAIN (BY SIMILARITY).
FT DISULFID 70 86 BY SIMILARITY.
FT DISULFID 168 244 BY SIMILARITY.
FT DISULFID 201 223 BY SIMILARITY.
FT DISULFID 234 262 BY SIMILARITY.
FT ACT_SITE 85 85 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 134 134 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 238 238 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT CARBOHYD 110 110 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 159 159 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 342 AA; 36729 MW; 0620DE88ED187D0F CRC64;

Query Match 32.3%; Score 631; DB 1; Length 342;
Best Local Similarity 41.4%; Pred. No. 5.2e-43;
Matches 146; Conservative 48; Mismatches 139; Indels 20; Gaps 10;

QY 1 MAQKGVLPGLGAVANDSYSLYGLVPSG-PARG-PPYCGGRP-BPSARIYGGNAQPGT 57
Db 1 MAPRVGLGLGLEAVT---ILLLGLLQSGIRADGTEASCGAVIQP--RITGGGSAKPGQ 55

QY 58 WPMQVSLHHGGGHCIGGSLIAPSWLSAAHCFMTNGTLEPAAEWSVLLGVHSDQGLDGA 117
Db 56 WPMQVSLTYDGNHVGCGSLVSNKWVSAHCFPREHSRE--AYEVKLGALHQLDSYNDT 112

QY 118 HTRAVAAIVVPANYSQVELGADLALRLASPSALGPVAVMPVCLPRASHRFVHGTACWATG 177
Db 113 VVHTVAQIITHSSYREEGSGQDIALIRLSSPVTFSSYIRPCLIPAANASFPNGLHCTVTG 172

QY 178 WGDVQADPLPLPWVLOVELRLLEATCCLYSQPGPNLTLIQLPGMLCAGYPEGRED 237
Db 173 WGHVAPSVLSQTPRPLQQLLEVLPLISRETCSCLYNINAVPEEPTIQDMLCAGYVKGKD 232

QY 238 TCGSGGGLPVCBEGGRWFOAGITSPFCGGRNRPVGTAVATYPAWIREQVMGSEPG 297
Db 233 ACQGSGLPSCPMEGIWYLAGIVSGDAGCAPNRPVGTAVTASTYASWIIHHV--AELQP 290

QY 298 APTQPKTKQSD---CLHQTAFLDNR--ILRLPLSHISVGVSTGTSLVLPWL 346
Db 291 RVVPQTQESQPDGHLCHNHPVFNAAAPKLLRVLPLGLTGLGLSL---WL 340

```

RESULT 3

```

PS88 RAT STANDARD; PRT; 342 AA.
AC Q9ESB7; Q9ER01;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Proctasin precursor (EC 3.4.21.-).
GN PRS8.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
CX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Kidney;
RA Adachi M., Kitamura K., Miyoshi T., Tomita K.;
RL Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA Wang C.;
RT "Molecular cloning and expression of rat proctasin.";
RL Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: Possesses a trypsin-like cleavage specificity (By
CC similarity).
CC -!- SUBUNIT: Heterodimer of two chains, light and heavy, held by a
CC disulfide bond (By similarity).
CC -!- SUBCELLULAR LOCATION: MEMBRANE-BOUND. SECRETED AFTER CLEAVAGE OF
CC ITS C-TERMINUS (BY SIMILARITY).

```

```

CC -!- SIMILARITY: Belongs to peptidase family S1.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AB017638; BAB20281.1; --
CC EMBL; AF202076; AAG32641.1; --
CC HSSP; P00734; IUVS.
CC
CC InterPro; IPR009003; Cys_Ser_trypsin.
CC InterPro; IPR001254; Peptidase_S1.
CC InterPro; IPR001314; Peptidase_S1A.
CC Pfam; PF00089; trypsin; 1.
CC PRINTS; PR00722; CHYMOTRYPSIN.
CC SMART; SM00020; Tryp_Spc; 1.
CC PROSITE; PS0240; TRYPSIN_DOM; 1.
CC PROSITE; PS00134; TRYPSIN_HIS; 1.
CC PROSITE; PS00135; TRYPSIN_SER; 1.
KW Hydrolase; Serine protease; Zymogen; Signal; Glycoprotein;
KW Transmembrane.
FT SIGNAL 1 29 POTENTIAL.
FT PROPEP 30 32 ACTIVATION PEPTIDE (BY SIMILARITY).
FT CHAIN 33 44 PROSTATIN LIGHT CHAIN.
FT CHAIN 45 322 PROSTATIN HEAVY CHAIN.
FT PROPEP 323 342 BY SIMILARITY.
FT TRANSMEM 320 340 POTENTIAL.
FT DOMAIN 45 286 SERINE PROTEASE.
FT DISULFID 37 154 INTERCHAIN (BY SIMILARITY).
FT DISULFID 70 86 BY SIMILARITY.
FT DISULFID 168 244 BY SIMILARITY.
FT DISULFID 201 223 BY SIMILARITY.
FT DISULFID 234 262 BY SIMILARITY.
FT ACT_SITE 85 85 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 134 134 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 238 238 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT CARBOHYD 159 159 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CONFLICT 17 17 I -> V (IN REF. 1).
FT CONFLICT 292 292 A -> V (IN REF. 1).
SQ SEQUENCE 342 AA; 36843 MW; 5ED1AF05D9213B98 CRC64;

Query Match 31.7%; Score 620; DB 1; Length 342;
Best Local Similarity 40.8%; Pred. No. 3.9e-42;
Matches 144; Conservative 50; Mismatches 139; Indels 20; Gaps 9;

QY 1 MAQKGVLPGLGAVANDSYSLYGLVPS--GPARGPPYCGRP-BPSARIYGGNAQPGT 57
Db 1 MALRVGLGLGLEALF---IDLLGLLQSGIRADGTEASCGAVIQP--RITGGGSAKPGQ 55

QY 58 WPMQVSLHHGGGHCIGGSLIAPSWLSAAHCFMTNGTLEPAAEWSVLLGVHSDQGLDGA 117
Db 56 WPMQVSLTYDGNHVGCGSLVSNKWVSAHCFPREHSKE---EYEVKLGALHQLDSFNDI 112

QY 118 HTRAVAAIVVPANYSQVELGADLALRLASPSALGPVAVMPVCLPRASHRFVHGTACWATG 177
Db 113 VVHTVAQIITHSSYREEGSGQDIALIRLSSPVTFSSYIRPCLIPAANASFPNGLHCTVTG 172

QY 178 WGDVQADPLPLPWVLOVELRLLEATCCLYSQPGPNLTLIQLPGMLCAGYPEGRED 237
Db 173 WGHVAPSVLSQTPRPLQQLLEVLPLISRETCSCLYNINAVPEEPTIQDMLCAGYVKGKD 232

QY 238 TCGSGGGLPVCBEGGRWFOAGITSPFCGGRNRPVGTAVATYPAWIREQVMGSEPG 297
Db 233 ACQGSGLPSCPMEGIWYLAGIVSGDAGCAPNRPVGTAVTASTYASWIIHHV--AELQP 290

QY 298 APTQPKTKQSD---CLHQTAF--LDSARILLRPLSHISVGVSTGTSLVLPWL 346
Db 291 RVVPQTQESQPDGHLCHNHPVFNAAAKLSRPLEPLSLTLGLFSL---WL 340

```



```

RESULT 4
TRYG HUMAN
ID TRYG HUMAN STANDARD; PRT; 321 AA.
AC Q9NR02; Q9C015; Q9NR08; Q9UBB2;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Trypsin gamma precursor (EC 3.4.21.-) (Transmembrane trypsinase).
GN TP5G1 OR TMT.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.
NCBI_TaxID=9606;
RX MEDLINE=20302813; PubMed=10843716;
RA Caughey G.H., Raymond W.W., Blount J.L., Hau L.W., Pallaoro M.,
RA Wolters P.J., Verghese G.M.;
RT "Characterization of human gamma-tryptases, novel members of the
RT chromosome 16p mast cell trypsinase and prostatic gene families.";
RT J. Immunol. 164:6566-6575 (2000).
RN [2]
RN SEQUENCE FROM N.A. (VARIANTS GAMMA-1 AND GAMMA-2).
RX MEDLINE=20302813; PubMed=10843716;
RA Caughey G.H., Raymond W.W., Blount J.L., Hau L.W., Pallaoro M.,
RA Wolters P.J., Verghese G.M.;
RT "Characterization of human gamma-tryptases, novel members of the
RT chromosome 16p mast cell trypsinase and prostatic gene families.";
RT J. Immunol. 164:6566-6575 (2000).
RN [2]
RN SEQUENCE FROM N.A.
RX MEDLINE=99452974; PubMed=10521469;
RA Wong G.W., Tang Y., Feyfant E., Sali A., Li Y., Huang C.,
RA Friend D.S., Krillis S.A., Stevens R.L.; trypsinase family of mouse and
RT "Identification of a new member of the trypsinase family of mouse and
RT human mast cell proteases which possesses a novel COOH-terminal
RT hydrophobic extension.";
RT J. Biol. Chem. 274:30784-30793 (1999).
RN [3]
RN SEQUENCE OF 220-321 FROM N.A.
RA Mittman S., Agnew W.S.;
RT "Organization and alternative splicing of CANALH.";
RL Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.
CC -!- SUBCELLULAR LOCATION: Membrane-anchored (Potential).
CC -!- TISSUE SPECIFICITY: Expressed in many tissues.
CC -!- POLYMORPHISM: There are two alleles; gamma-I and gamma-II which
CC differ by 5 residues.
CC -!- SIMILARITY: Belongs to peptidase family S1. Trypsin subfamily.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
CC EMBL; AF191031; AAF76457.1; -
CC EMBL; AF195508; AAF76458.1; -
CC EMBL; AF175759; AAF03697.1; -
CC EMBL; AF175522; AAF03695.1; -
CC EMBL; AF223563; AAG48852.2; -
CC HSP; P00763; 1DPO.
CC MEROPS; S01.028; -.
CC Genew; HGNC:14134; TP5G1.
CC InterPro; IPR009003; Cys_Ser_trypsin.
CC InterPro; IPR001254; Peptidase_S1.
CC InterPro; IPR001314; Peptidase_S1A.
CC Pfam; PF00089; trypsin, 1.
CC PRINTS; PR00722; CHYMOTRYPSIN.
CC PROSITE; PS02240; TRYPsin DOM; 1.
CC PROSITE; PS00134; TRYPsin_HIS; 1.
CC PROSITE; PS00135; TRYPsin_SER; FALSE NEG.
CC KW Hydrolyase; Serine protease; Signal; Glycoprotein; Zymogen;
KW Transmembrane; Polymorphism.
FT SIGNAL 1 19 POTENTIAL.
FT CHAIN 20 36 TRYPTASE GAMMA LIGHT CHAIN.
FT CHAIN 38 321 TRYPTASE GAMMA HEAVY CHAIN.
FT TRANSMEM 284 304 POTENTIAL.
FT ACT_SITE 78 78 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 125 125 CHARGE RELAY SYSTEM (BY SIMILARITY).

```

```

FT ACT_SITE 222 222 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT DISULFID 26 145 INTERCHAIN (POTENTIAL).
FT DISULFID 63 79 BY SIMILARITY.
FT DISULFID 159 228 BY SIMILARITY.
FT DISULFID 192 210 BY SIMILARITY.
FT DISULFID 218 246 BY SIMILARITY.
FT CARBOHYD 85 85 M-LINKED (GLCNAC...) (POTENTIAL).
FT VARIANT 60 60 M -> V (IN GAMMA-II).
FT VARIANT 126 126 /FTID=VAR 012097.
FT VARIANT 132 132 I -> M (IN GAMMA-II).
FT VARIANT 132 132 /FTID=VAR 012098.
FT VARIANT 204 204 S -> T (IN GAMMA-II).
FT VARIANT 204 204 /FTID=VAR 012099.
FT VARIANT 288 288 L -> I (IN GAMMA-II).
FT VARIANT 288 288 /FTID=VAR 012100.
FT VARIANT 288 288 L -> F (IN GAMMA-II).
FT CONFLICT 160 160 /FTID=VAR 012101.
FT CONFLICT 321 321 W -> S (IN REF. 1).
SQ SEQUENCE 321 AA; 33827 MW; FFF7B06E3C4A962D CRC64;

Query Match 29.8%; Score 582; DB 1; Length 321;
Best Local Similarity 46.0%; Pred. No. 3.7e-39;
Matches 128; Conservative 35; Mismatches 87; Indels 28; Gaps 11;

QY 36 PYCGRPEPS---ARIYVGSNAQPTWPMQVSIHHGGHICGSLIAPSWVLSAAHCFWTN 92
Db 24 PCGGRPQVSDAGRIYVGGHAPAGAPWQASLIRLRMRHVCGSLLSPQWLTAHCF--S 81
QY 93 GTLEPAASVLLGVHSDGDL---GAHTRAVAAIVVPANYS-QVELGADIALRLASP 148
Db 82 GSLN-SSDYQVHL-----GELEITLSPHFSTVRQIIHSSPSGGQGTSGDIALVELSVP 134
QY 149 ASLGPVAVWPVCLPRASHRFVHGTCACWATGWDVQVADPLPLPWLOEVELRLGEATQC 208
Db 135 VTLSRIILPVCLPEASDDFCGIRCWVIGWYREGEPLPPYSLSREKVSVDVETCR 194
QY 209 LYSQGPENLTLQIIPGMLCAGYPERDRDTCQDGGGLVCEGGRWFQAGITSFGCG 268
Db 195 DY--FPGGSIQ--FDMLCA---RPGDGCDDSGGLVCGVQVAGVSWGEGCG 247
QY 269 RNRPGVTAATYAWIREQVW---GSEPGPAFTQP 303
Db 248 RNRPGVTRPAYVWIRRHITASGSESG--YPLP 283

RESULT 5
PR27 HUMAN STANDARD; PRT; 290 AA.
ID PR27 HUMAN
AC Q9BQR3;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Pancreas precursor (EC 3.4.21.-) (Marapsin) (Channel-activating
DE protease 2) (CAPH2).
GN PRSS27 OR MEN.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.
NCBI_TaxID=9606;
RX [1]
RX SEQUENCE FROM N.A.
RA Fortunato M., Dando P.M., Rawlings N.D., Barrett A.J.;
RT "Cloning, sequencing and expression of marapsin, a human serine
RT proteinase.";
RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.
RN [2]
RN SEQUENCE FROM N.A.
RP Okaze H., Hayashi A., Kozuma S., Saito T.;
RA Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
RN [3]
RN SEQUENCE FROM N.A.
RX MEDLINE=22439795; PubMed=12441343;
RA Bhagwandin V.J., Hau L.W.-T., Mallen-St Clair J., Wolters P.J.,

```

Caughey G.H.;
 "Structure and activity of human pancreas, a novel tryptic serine
 peptidase expressed primarily by the pancreas.";
 J. Biol. Chem. 278:3363-3371(2003).
 CC -1- SUBCELLULAR LOCATION: Secreted.
 CC -1- TISSUE SPECIFICITY: Expressed predominantly in the pancreas.
 CC -1- PTM: N-glycosylated.
 CC -1- SIMILARITY: Belongs to peptidase family S1.

 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (see <http://www.isb-sib.ch/announce/>)
 or send an email to license@sib-sib.ch.

 DR EMBL; AJ306593; CAC35467.1; -;
 DR EMBL; AB056161; BAB85497.1; -;
 DR EMBL; AY030095; AAX38168.1; -;
 DR HSSP; P00734; 1UUS.
 DR MEROPS; S01.074; -;
 DR Genew; HGNC:15475; PRS927.
 DR MIM; 608018; -;
 DR InterPro; IPR009003; Cys Ser trypsin.
 DR InterPro; IPR001254; Peptidase S1.
 DR InterPro; IPR001314; Peptidase_S1A.
 DR Pfam; PF00089; trypsin; 1.
 DR PRINTS; PR00722; CHYMOTRYPSIN.
 DR SMART; SM0020; TRYPSIN; 1.
 DR PROSITE; PS00240; TRYPSIN_DOM; 1.
 DR PROSITE; PS00134; TRYPSIN_HIS; 1.
 DR PROSITE; PS00135; TRYPSIN_SER; 1.
 DR Hydrolase; Serine protease; Zymogen; Signal; Glycoprotein.
 FT SIGNAL 1 22 POTENTIAL.
 FT PROPEP 23 34 POTENTIAL.
 FT CHAIN 35 290 PANCREASIN.
 FT DOMAIN 35 277 SERINE PROTEASE.
 FT ACT_SITE 75 75 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT ACT_SITE 124 124 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT ACT_SITE 229 229 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT DISULFID 60 76 BY SIMILARITY.
 FT DISULFID 158 235 BY SIMILARITY.
 FT DISULFID 191 214 BY SIMILARITY.
 FT DISULFID 225 253 BY SIMILARITY.
 FT CARBOHYD 55 55 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 79 79 N-LINKED (GLCNAC...) (POTENTIAL).
 SQ SEQUENCE 290 AA; 67BDC93EC70BFF7B CRC64;
 Query Match 28.8%; Score 563; DB 1; Length 290;
 Best Local Similarity 42.2%; Pred. No. 1.1e-37;
 Matches 117; Conservative 32; Mismatches 114; Indels 14; Gaps 6;
 QY 32 ARGPPYCRPEPARIYVGSNAQPTWQVSLHGGHICGSLIAPSWVLSAAHCFMT 91
 Db 20 AKAATACGRPMNLNRMVGGQDQEGEWPMQVSIQRNGSHFCGSLIACQWLVTAHCP-- 77
 QY 92 NGTLEPAAEWSVLLGVHSGDGLDGAHTRAAVAVIPAN--YSQVELGADLALLASPA 149
 Db 78 NNTSETSL-YQVLLGARQLVQ--GPHAMVAVRVQSNPLVQGTASSADVALVEAPV 134
 QY 150 SUGPAWVPVCLPRASHRFVGHGTACWATGVDVQVADPLPLPWVQVVELLLGEATCOQL 209
 Db 135 PFTNVILPVCLPDPSPVIFETGMNVCWVWGSPSEEDLPEPRILQKLAVPIDTPKCNLL 194
 QY 210 YSQPGFNLTQLLPQ-MLCAGVPEGRDTCGGDGGPLVCEGGVWFOAGITTSFGGCG 268
 Db 195 YSKOTFEGYQPTIKNDMLCAGFEKGKDACGSGGGLVCLVGVSLQAGVISMWGCA 254
 QY 269 RNRNPGVATVATYEAWIREQVWVGSPGPPAFPTQPOK 305
 Db 255 RNRNPGVYRTVAHNWIIHRII-----PKLQFQPAR 285

RESULT 6
 TMS6 HUMAN STANDARD; PRT; 811 AA.
 ID Q8IU80; Q8IU82; Q8IXV8;
 DT 10-OCT-2003 (Rel. 42, Created)
 DT 10-OCT-2003 (Rel. 42, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Transmembrane protease, serine 6 (EC 3.4.21.-) (Matriptase-2).
 GN TMPRSS6.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A. (ISOFORM 1).
 RA Hooper J.D., Quigley J.P.;
 RT "TMPRSS6, a new type II transmembrane serine protease.";
 RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RA MEDLINE=20057165; PubMed=10591208;
 RA Dunham I., Hunt A.R., Collins J.E., Bruskiewicz R., Beare D.M.,
 RA Clump M., Smink L.J., Ainscough R., Almeida J.P., Babbage A.K.,
 RA Bagguley C., Bailey J., Barlow K.F., Bates K.N., Beasley O.P.,
 RA Bird C.P., Blakey S.E., Bridgeman A.M., Buck D., Burgess J.,
 RA Burrill W.D., Burton J., Carder C., Carter N.P., Chen Y., Clark G.,
 RA Clegg S.M., Cobley V.E., Cole C.G., Collier R.E., Connor R.,
 RA Conroy D., Corby N.R., Coville G.J., Cox A.V., Davis J., Dawson E.,
 RA Dhani P.D., Dockree C., Dodsworth S.J., Durbin R.M., Ellington A.G.,
 RA Evans K.L., Fey J.M., Fleming K., French L., Garner A.A.,
 RA Gilbert J.G.R., Goward M.E., Grafham D.V., Griffiths M.N.D., Hall C.,
 RA Hall R.E., Hall-Tamlyn G., Heathcote R.W., Ho S., Holmes S.,
 RA Hunt S.E., Jones M.C., Kershaw J., Kimberley A.M., King A.,
 RA Laird G.K., Langford C.F., Leversha M.A., Lloyd C., Lloyd D.M.,
 RA Martyn I.D., Mashreghi-Mohammadi M., Matthews L.H., Mccann O.T.,
 RA Mcclay J., McLaren S., McMurray A.A., Milne S.A., Mortimore B.J.,
 RA Odell C.N., Pavitt R., Pearce A.V., Pearson D., Phillimore B.J.C.T.,
 RA Phillips S.H., Plumb R.W., Ramsay H., Ramsey Y., Rogers L., Ross M.T.,
 RA Scott C.E., Sehra H.K., Skuce C.D., Smalley S., Smith M.L.,
 RA Soderlund C., Spragon L., Steward C.A., Sulston J.E., Swann R.M.,
 RA Vaudin M., Wall M., Wallis J.M., Whiteley M.N., Willey D.L.,
 RA Williams L., Williams S.A., Williamson H., Wilmer T.E., Wilming L.,
 RA Wright C.L., Hubbard T., Bentley D.R., Beck S., Rogers J., Shimizu N.,
 RA Minoshima S., Kawasaki K., Sasaki T., Asakawa S., Kudon J.,
 RA Shintani A., Shibuya K., Yoshizaki Y., Aoki N., Mitsuyama S.,
 RA Roe B.A., Chen F., Chu L., Crabtree J., Deschamps S., Do A., Do T.,
 RA Dorman A., Fang F., Fu Y., Hu P., Hua A., Kenton S., Lai H., Lao H.I.,
 RA Lewis J., Lewis S., Lin S.-P., Loh P., Mala J.E., Nguyen T., Pan H.,
 RA Phan S., Qi S., Qian Y., Ray L., Ren Q., Shaull S., Sloan D., Song L.,
 RA Wang Q., Wang Z., White J., Willingham D., Wu H., Yao Z.,
 RA Zhan M., Zhang G., Chisoe S., Murray J., Miller N., Mink P.,
 RA Fulton R., Johnson D., Benis G., Bentley D., Bradshaw H., Bourne S.,
 RA Cordes M., Du Z., Fulton L., Goela D., Graves T., Hawkins J.,
 RA Hinds K., Kemp K., Latreille P., Layman D., Ozersky P., Rohlfing T.,
 RA Scheet P., Walker C., Wamsley A., Wohlmann P., Pepin K., Nelson J.,
 RA Korfi I., Bedal J.A., Hillier L.W., Mardis E., Waterston R.,
 RA Wilson R., Emanuel B.S., Shaikh T., Kurahashi H., Saitta S.,
 RA Budarf M.I., Mcdermid H.E., Johnson A., Wong A.C.C., Morrow B.E.,
 RA Edelmann L., Kim U.J., Shizuya H., Simon M.I., Dumanski J.P.,
 RA Peyrard M., Kedra D., Seroussi E., Fransson I., Tapia I., Bruder C.E.,
 RA O'Brien K.P., Wilkinson P., Bodenteich A., Hartman K., Hu X.,
 RA Khan A.S., Lane L., Tilahun Y., Wright H.,
 RT "The DNA sequence of human chromosome 22.";
 RL Nature 402:489-495(1999).
 RN [3]
 RP SEQUENCE FROM N.A. (ISOFORM 2).
 RC TISSUE=Brain;
 RX MEDLINE=2238257; PubMed=12477932;
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.F., Schuler G.D.,
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.M., Bhat N.K.,
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,

RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
 RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
 RA Raha S.S., Loquellano N.A., Peters G.J., Aramson R.D., Mullany S.J.,
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
 RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 RA Fahey J., Helton E., Kettman M., Madan A., Rodrigues S., Sanchez A.,
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
 RA Butterfield Y.S.N., Krzyzinski M.I., Skalska U., Smailus D.B.,
 RA Sutterch A., Schein J.E., Jones S.J.M., Marra M.A.,
 PT "Generation and initial analysis of more than 15,000 full-length
 RT human and mouse cDNA sequences."
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
 RN [4]
 RP SEQUENCE OF 9-811 FROM N.A. (ISOFORM 1), AND TISSUE SPECIFICITY.
 RC TISSUE=Fetal liver;
 RX MEDLINE=22241917; PubMed=12149247;
 RA Velasco G., Cal S., Quesada V., Sanchez L.M., Lopez-Otin C.;
 RT "Matritase-2, a membrane-bound mosaic serine proteinase predominantly
 RT expressed in human liver and showing degrading activity against
 RT extracellular matrix proteins."
 RL J. Biol. Chem. 277:37637-37646(2002).
 RP [5]
 RP REVIEW
 RX MEDLINE=22668120; PubMed=12784999;
 RA Netzel-Arnett S., Hooper J.D., Szabo R., Madison E.L., Quigley J.P.,
 RA Bugge T.H., Antalis T.M.;
 RT "Membrane anchored serine proteases: a rapidly expanding group of cell
 RT surface proteolytic enzymes with potential roles in cancer."
 RL Cancer Metastasis Rev. 22:237-258(2003).
 CC -!- FUNCTION: May play a specialized role in matrix remodeling
 CC processes in liver.
 CC -!- SUBCELLULAR LOCATION: Type II membrane protein.
 CC -!- ALTERNATIVE PRODUCTS:
 CC Event=Alternative splicing; Named isoforms=2;
 CC Name=1;
 CC IsoId=Q8IU80-1; Sequence=Displayed;
 CC Name=2;
 CC IsoId=Q8IU80-2; Sequence=VSP_008379, VSP_008380;
 CC Note=No experimental confirmation available;
 CC TISSUE SPECIFICITY: Liver specific.
 CC -!- SIMILARITY: Belongs to peptidase family S1.
 CC -!- SIMILARITY: Contains 2 CUB domains.
 CC -!- SIMILARITY: Contains 3 LDL-receptor class A domains.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; AJ319876; CAC85953.1; ALT_INIT.
 CC EMBL; AY055383; AL16413.1; -
 CC EMBL; AY055384; AL16414.1; -
 CC EMBL; AL022314; -; NOT ANNOTATED CDS.
 CC EMBL; BC039082; AH39082.1; -
 CC InterPro; IPR000859; CUB.
 CC InterPro; IPR009003; Cys Ser trypsin.
 CC InterPro; IPR002172; LDL receptor A.
 CC InterPro; IPR001254; Peptidase_S1.
 CC InterPro; IPR001314; Peptidase_S1A.
 CC Pfam; PF00431; CUB; 1.
 CC Pfam; PF00057; ldl_recept_a; 2.
 CC Pfam; PF00089; trypsin; 1.
 CC PRINTS; PR00722; CHYMOTRYPSIN.
 CC PRINTS; PR00261; LDLRECEPTOR.
 CC SMART; SM00192; LDLa; 3.
 CC SMART; SM00020; Tryp_Spc; 1.

DR PROSITE; PS01180; CUB; 1.
 DR PROSITE; PS01209; LDLRA_1; 1.
 DR PROSITE; PS00068; LDLRA_2; 3.
 DR PROSITE; PS00240; TRYPSIN_DOM; 1.
 DR PROSITE; PS00134; TRYPSIN_HIS; 1.
 DR PROSITE; PS00135; TRYPSIN_SER; 1.
 KW Hydrolase; Serine protease; Repeat; Signal-anchor; Transmembrane;
 KW Glycoprotein; Alternative splicing.
 FT DOMAIN 1 55 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 56 76 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
 FT (POTENTIAL).
 FT (EXTRACELLULAR (POTENTIAL)).
 FT CUB 1.
 FT CUB 2.
 FT LDL-RECEPTOR CLASS A 1.
 FT LDL-RECEPTOR CLASS A 2.
 FT LDL-RECEPTOR CLASS A 3.
 FT SERINE PROTEASE.
 FT CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT ACT_SITE 617 617 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT ACT_SITE 668 668 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT ACT_SITE 762 762 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT CARBOHYD 136 136 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 184 184 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 216 216 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 338 338 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 433 433 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 453 453 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 518 518 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT VARSPLIC 409 461 LCGLRIILQPYAERIPVAVATAGITINFTQSILTCGVRVHY
 FT SLEAQVFNLRGASRGSGWGCQACCP (in isoform
 FT 2).
 FT /FTId=VSP_008379.
 FT Missing (in isoform 2).
 FT /FTId=VSP_008380.
 FT A -> V (IN REF. 4).
 FT CONFLICT 116 116
 FT SEQUENCE 811 AA; 89999 MW; 7EEF193F655DDE9D CRC64;
 SQ
 Query Match 28.0%; Score 547; DB 1; Length 811;
 Best Local Similarity 43.3%; Pred. No. 6.2e-36;
 Matches 110; Conservative 42; Mismatches 90; Indels 12; Gaps 4;
 QY 38 CGREPSRIYVGSNAQPGTWQVSLHHGGHICGSLIAPSWVLAAHCFWINGTLEP 97
 DB 568 CGLQGPSRIYVGSNAQPGTWQVSLHHGGHICGSLIAPSWVLAAHCFWINGTLEP 626
 QY 98 AAEMSVLLGVHSQDGLDGAHTRAAVAVPVPANYSQVELGADLALLRLASPASLGPAVWP 157
 DB 627 TVLWTVFLGKYNQSRWPFGEVSFKVSRLLHPHYEEDSHDYDVALQLDHPVRSAAVRP 686
 QY 158 VCLPRASHRFVHGACWATGWDQVQADPLPWLQVELLGLGEATCCCLYSQPGPN 217
 DB 687 VCLPARSHFFPEGLHCWITGWALEGGPISN--ALQKVDVQLIPQDLCSVEY----- 737
 QY 218 LTLQLLQMLCAGYPEGRRDTCQSGGGLVCEE--GGRWFQAGITSFGCGRRNRPGVF 276
 DB 738 -RYQVTPRMLCAGYKKGKADCCQSGGGLVCKALSGRFLAGLVSGLGCGRNVEGVY 796
 QY 277 TAVATYEAWIREQV 290
 DB 797 TRITGVISWIIQVQV 810
 RESULT 7
 BSS4 HUMAN STANDARD; PRT; 317 AA.
 ID BSS4 HUMAN
 AC Q9GZ4; C43342;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Brain-specific serine protease 4 precursor (EC 3.4.21.-) (BSSP-4)
 DE (SP001LA).
 GN PRSS22 OR PRSS26 OR BSSP4.

RC TISSUE-Prostate;
RA Yamaguchi N., Mitsui S.;
RT "Molecular cloning of a novel transmembrane serine protease expressed
in human prostate.";
RL Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RA Tanimoto H., Underwood L.J., Wang Y., Shigenasa K., Farnley T.H.,
RA O'Brien T.J.;
RL Submitted (APR-1998) to the EMBL/GenBank/DBJ databases.
RN [5]
RP SEQUENCE FROM N.A.
RA TISSUE-Blood, and Muscle;
RX MEDLINE=2238257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grinwood J., Schmutz J., Myers R.M.,
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smalilus D.E.,
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length
human and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
RN [6]
RP SEQUENCE OF 340-664 FROM N.A.
RA Cao J., Fan W., Zheng S.;
RT "Genomic analysis of a novel human serine protease SN19.";
RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
RN [7]
RP CHARACTERIZATION.
RC TISSUE-Milk;
RX MEDLINE=99303582; PubMed=10373425;
RA Lin C.Y., Anders J., Johnson M., Dickson R.B.;
RT "Purification and characterization of a complex containing matriptase
and a Kunitz-type serine protease inhibitor from human milk.";
RL J. Biol. Chem. 274:18237-18242 (1999).
CC -!- FUNCTION: Degrades extracellular matrix. Proposed to play a role
in breast cancer invasion and metastasis. Exhibits trypsin-like
activity as defined by cleavage of synthetic substrates with Arg
or Lys as the P1 site.
CC -!- SUBCELLULAR LOCATION: Type II membrane protein (Probable).
CC -!- SIMILARITY: Belongs to peptidase family S1.
CC -!- SIMILARITY: Contains 2 CUB domains.
CC -!- SIMILARITY: Contains 4 LDL-receptor class A domains.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@sib-sib.ch).
CC
CC EMBL; AF118224; AAD42765.2; -
CC EMBL; AF133086; AAF00109.1; -
CC EMBL; AF030036; BAB20376.1; -
CC EMBL; AF057145; AAG15395.1; -
CC EMBL; BC005826; AAH05826.1; -
CC EMBL; BC030532; AAH30532.1; -
CC EMBL; AF283256; AAG13949.1; -
CC HSSP; P00763; LDPO.
CC Genew; HGNC:11344; ST14.

DR MM; 606797; -
DR MEROPS; S01.302; -
DR GO; GO:0005887; F: integral to plasma membrane; TAS.
DR GO; GO:0008236; F: serine-type peptidase activity; TAS.
DR GO; GO:0006508; P: proteolysis and peptidolysis; TAS.
DR InterPro; IPR000859; CUB.
DR InterPro; IPR009003; Cys Ser trypsin.
DR InterPro; IPR002172; LDL_receptor_A.
DR InterPro; IPR001254; Peptidase_S1.
DR InterPro; IPR001314; Peptidase_S1A.
DR Pfam; PF00431; CUB; 2.
DR Pfam; PF00057; ldl_recept_a; 4.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PR00722; CHYMOTRYPSIN.
DR PRINTS; PR00261; LDLRECEPTOR.
DR SMART; SM00042; CUB; 2.
DR SMART; SM00192; LDLa; 3.
DR SMART; SM00020; Tryp_Spc; 1.
DR PROSITE; PS01180; CUB; 2.
DR PROSITE; PS01209; LDLRA_1; 2.
DR PROSITE; PS00068; LDLRA_2; 4.
DR PROSITE; PS50240; TRYPsin_DOM; 1.
DR PROSITE; PS00134; TRYPsin_HIS; 1.
DR PROSITE; PS00135; TRYPsin_SER; 1.
KW Signal-anchor; Glycoprotein; Hydrolase; Serine protease;
KW Transmembrane; Repeat.
FT DOMAIN 1 55
FT TRANSMEM 56 76
FT
FT DOMAIN 77 855
FT DOMAIN 214 334
FT DOMAIN 340 447
FT DOMAIN 452 487
FT DOMAIN 487 524
FT DOMAIN 524 560
FT DOMAIN 566 603
FT DOMAIN 615 854
FT ACT_SITE 656 656
FT ACT_SITE 711 711
FT ACT_SITE 805 805
FT CARBOHYD 109 109
FT CARBOHYD 302 302
FT CARBOHYD 485 485
FT CARBOHYD 772 772
FT CONFLICT 327 329
FT CONFLICT 381 381
FT CONFLICT 674 674
FT CONFLICT 855 855
SQ SEQUENCE 855 AA; 94769 MW; 26143132C01F99C9 CRC64;
Query Match 27.6%; Score 540; DB 1; Length 855;
Best Local Similarity 40.2%; Pred. No. 2.4e-35;
Matches 113; Conservative 39; Mismatches 95; Indels 34; Gaps 7;
QY 32 ARGPPYCGRPEP-----SARIVGGSNAQPGTWPMQVSLHH-GGGHI 71
DB 581 SKGNPECDGKDCSDGSDKDCGLRSFTQARVVGTDADGEGWPMQVSLHAGQGGHI 640
QY 72 CGGSLIAPSWLVAHACFMING--TLEPAEWSVLLGVHSQ-DGPLDGAHTRAVAAIVVP 128
DB 641 CGASLISFNWLSVAHCHYIDDRGFRYSPTQTWAFGLHDQSQRSAFGVQERRLKRILSH 700
QY 129 ANYSQVEIGADLALRLASPASLGAFAVWFCVLPRAASHFVHGTCATGAGWDVQADPLP 188
DB 701 PFENFDFTDYDIALLELEKPAEYSWVRPICLPDASHVFPAGKAIWVTGWGHTQGGTGA 760
QY 189 LPWLQVEVLLGLGEATCCQLYSGPGPNLTILPGMLCAGYPERDRDTCQDSDGGPL- 247
DB 761 L--ILQGEIRVINGTTCENLLPQ-----QITPRMCMVGFSGGVDSCQDSDGGPLS 810
QY 248 VCEGGRWFOAGITSGFGCGCRNRNRPVGTAVATYEAWARE 288
DB 811 SVEADGRIFQAGVSWGDCGCAQRNKPQVYTRLPFRDWIKE 851


```

QY 129 ANYSQVELGADLALLRLASPASLGAFAVWPVCLPRASHRFVHGTACWATGSDVOEADPLP 188
Db 701 PSFNDFTFDYDIALLELEKSEVEYTVVRPICLPDATHVFPAGKAIWYTGWHTKEGCTGA 760
QY 189 LPWLQVEVRLLEGEATCCQLYSQPGFENLTQLPGLCAGYEGEGRDTCQDGGGFL- 247
Db 761 L--ILQGEIRVINGTTCEDLMPQ-----QITPRMCMGVFISGGVDSQCGSGGFLS 810
QY 248 VCEGGRFQAGITSFGCGRRNRPGVFTAVATYEAWARE 288
Db 811 SAEDGRMFQAGVSWGEGCAQRNKPQVYTRLPVVRDRIKE 851

RESULT 14
BSS4 MOUSE
ID BSS4 MOUSE STANDARD; PRT; 306 AA.
AC O9ER10;
DT 16-OCT-2001 (Rel. 40, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DE Brain-specific serine protease 4 precursor (EC 3.4.21.-) (BSSP-4).
GN PRS222 OR PRS26 OR BSSP4.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RA Mitsui S., Okui A., Kominami K., Yamaguchi N.;
RT "Cloning and characterization of a novel serine protease, mBSSP-4.";
RL Submitted (JAN-1998) to the EMBL/GenBank/DBSJ databases.
CC -!- SIMILARITY: Belongs to peptidase family S1.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AB010778; BAB20262.1; -.
CC HSSP; P00763; IDPO.
CC MEROPS; S01.252; -.
CC MGD; MGI:1918085; 4733401N09Rik.
CC InterPro; IPR009003; Cys Ser trypsin.
CC InterPro; IPR001254; Peptidase_S1.
CC InterPro; IPR001314; Peptidase_S1A.
CC Pfam; PF00089; trypsin; 1.
CC PRINTS; PR00722; CHYMOTRYPSIN.
CC SMART; SM00020; Tryp_SPC; 1.
CC PROSITE; PS0240; TRYPSIN_DOM; 1.
CC PROSITE; PS00134; TRYPSIN_HIS; 1.
CC PROSITE; PS00135; TRYPSIN_SER; 1.
KW Hydrolase; Serine protease; Signal.
FT SIGNAL 1 32
FT CHAIN 33 306 BRAIN-SPECIFIC SERINE PROTEASE 4.
FT ACT_SITE 90 90 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 141 141 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 242 242 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT DISULFID 75 91 BY SIMILARITY.
FT DISULFID 175 248 BY SIMILARITY.
FT DISULFID 208 227 BY SIMILARITY.
FT DISULFID 238 266 BY SIMILARITY.
FT CARBOHYD 70 70 N-LINKED (GLCNAC...) (POTENTIAL).
SQ SEQUENCE 306 AA; 33262 MW; FBBF03C0C285E7E8 CRC64;

Query Match
Best Local Similarity 37.1%; Pred. No. 6.2e-34;
Matches 109; Conservative 47; Mismatches 118; Indels 20; Gaps 6;

```

```

QY 9 PGOLGAVANSDSYSLYGL-----VSPGARGPPYCGRPSPSARIIVGSGNAQFQTPW 59
Db 7 PPALG-----GDQFSILLVLLTSTAPISATIRVSPDCGKQPOLNRIVGGEDSMDAQP 62
QY 60 WQVSLHHGGGHI CGSLLIAPSWSLAAHCFMTNGTLEPAEWSVILGVHSQDGLDGAHT 119
Db 63 WIVSILLKNGSHHCAGSLTNRWVYTAHCFKSN--MDKPSLFFVLLGAWKLGGSPRSQK 120
QY 120 RAVAAIVVPANYSOVE--LGADLALLRLASPASLGAFAVWPVCLPRASHRFVHGTACWATGW 178
Db 121 VGTAWLPHPRYSWKEGTHADIALVLEHSIQSERILPICLPDSSVRLPKPKDCWIAGW 180
QY 179 GDVQEADPLPWLQVEVRLLEGEATCCQLYSQPGFENLTQLPGLCAGYEGEGRDTC 238
Db 181 GSIQDGVLPHPQTLQKLVPIIDSELCKSLYWRGAGQE---AITEGMLCAGYLEGRDA 237
QY 239 CQDSSGGPLVCEGGRWFQAGITSFGCGRRNRPGVFTAVATYEAWAREVQVMG 292
Db 238 CLGDSGGPLMCQVDHLLTGLTIISWGEGCA--DRPGVYTSLLAHRSWVQRIVQG 290

RESULT 15
TRYT CANFA
ID TRYT CANFA STANDARD; PRT; 275 AA.
AC P15944;
DT 01-APR-1990 (Rel. 14, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Tryptase precursor (EC 3.4.21.59).
OS Canis familiaris (Dog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
OX NCBI_TaxID=9615;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=89352460; PubMed=2504277;
RA Vanderslice P., Craik C.S., Nadel J.A., Caughey G.H.;
RT "Molecular cloning of dog mast cell tryptase and a related protease:
RT structural evidence of a unique mode of serine protease activation.";
RL Biochemistry 28:4148-4155(1989).
CC -!- FUNCTION: Tryptase is the major neutral protease present in mast
CC cells and is secreted upon the coupled activation-degranulation
CC response of this cell type.
CC -!- CATALYTIC ACTIVITY: Preferential cleavage: Arg-|-, Lys-|-, but
CC with more restricted specificity than trypsin.
CC -!- SUBUNIT: Homotetramer.
CC -!- SUBCELLULAR LOCATION: Released from the secretory granules upon
CC mast cell activation.
CC -!- SIMILARITY: Belongs to peptidase family S1. Tryptase subfamily.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M24664; AAA30854.1; -.
CC FIR; A32410; A32410.
CC HSSP; P20231; IAAO.
CC MEROPS; S01.143; -.
CC InterPro; IPR009003; Cys Ser trypsin.
CC InterPro; IPR001254; Peptidase_S1.
CC InterPro; IPR001314; Peptidase_S1A.
CC Pfam; PF00089; trypsin; 1.
CC PRINTS; PR00722; CHYMOTRYPSIN.
CC SMART; SM00020; Tryp_SPC; 1.
CC PROSITE; PS0240; TRYPSIN_DOM; 1.
CC PROSITE; PS00134; TRYPSIN_HIS; 1.
CC PROSITE; PS00135; TRYPSIN_SER; 1.
KW Hydrolase; Serine protease; Signal; Glycoprotein; Zymogen.
FT SIGNAL 1 20
FT SIGNAL 1 20

```

FT	PROPEP	21	30	ACTIVATION PEPTIDE (BY SIMILARITY).
FT	CHAIN	31	275	TRYPTASE.
FT	ACT_SITE	74	74	CHARGE RELAY SYSTEM (BY SIMILARITY).
FT	ACT_SITE	121	121	CHARGE RELAY SYSTEM (BY SIMILARITY).
FT	ACT_SITE	224	224	CHARGE RELAY SYSTEM (BY SIMILARITY).
FT	DISULFID	59	75	BY SIMILARITY.
FT	DISULFID	155	230	BY SIMILARITY.
FT	DISULFID	188	211	BY SIMILARITY.
FT	DISULFID	220	248	BY SIMILARITY.
FT	CARBOHYD	132	132	N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ	SEQUENCE	275 AA;	30088 MW;	C3B869251F24BD58 CRC64;
Query Match 26.3%; Score 514.5; DB 1; Length 275;				
Best Local Similarity 42.2%; Pred. No. 7.2e-34;				
Matches 124; Conservative 27; Mismatches 100; Indels 43; Gaps 9;				
QY	23	LYGLVPSGARGPPYCGRPEPSARI	VCGSNAQPTWQVSLHHGG---GHICGGSLLAP	79
Db	12	LGSLLVPVSPAP-----GQALQR	VGIVGREGPKWPKVQVSLRLKGQYWRHICGGSLLHP	66
QY	80	SWVLSAAHCFMTNGTLEPAAEWSVLL	---GVHSQDGLDGAHTRAVAAIWPANYSQVEL	136
Db	67	QNVLTAAHCVGN--VVCPEETRVQLR	EQHLYQD-----HLLPNRIVMHPNYTPEN	118
QY	137	GADLALLRLASPASLGPVAVPVCL	PRASHRFVHGTAQWAGDVQOEAADPLPWVLQEV	196
Db	119	GADIALLELEDPVNVSAAHVQVPTL	PPALQTFPTGTCWVTGWDVHSGTLPPLPPFLKQV	178
QY	197	ELRLIGEATCQCLYSQGFENLTQL	ILPG-----MLCAGYPEGRDTCQDSSGGPLVC	249
Db	179	KVPIVENSMDVQY-----HLGL	STGDGVRIVREDMLCAG--NSKSDSCQDSSGGPLVC	230
QY	250	BEGGRWFQAGITSFGFCGRRNR	PGVFTAVATYEAWIREQVMGSEPGPAFTQP	303
Db	231	RVRGWLQAGVVSWEGBGCAQPN	RPGIYTRVAYYLDWIHQYV-----PKEP	275

Search completed: February 25, 2004, 15:47:56
Job time : 52 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 25, 2004, 15:36:03 ; Search time 84 Seconds
(without alignments)
1340.952 Million cell updates/sec

Title: US-10-037-417-46

Perfect score: 1953

Sequence: 1 MAQKGLGPGQLGAVANSDS.....TKSLVLPWLSPHSLGLGWGF 357

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL_25:*
1: sp archaea:*
2: sp bacteria:*
3: sp fungi:*
4: sp human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phase:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rvirus:*
16: sp_bacteriap:*
17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1202	61.5	766	Q8NBY4	Q8NBY4 homo sapien
2	630.5	32.3	339	Q991A4	Q991A4 mus musculus
3	613	31.4	340	Q8BJV6	Q8BJV6 mus musculus
4	603	30.9	277	Q80WM7	Q80WM7 mus musculus
5	603	30.9	331	Q80X17	Q80X17 mus musculus
6	600	30.7	331	Q8RIA6	Q8RIA6 mus musculus
7	590	30.2	284	Q8NF86	Q8NF86 homo sapien
8	590	30.2	327	Q8N171	Q8N171 homo sapien
9	582	29.8	321	Q96RZ8	Q96RZ8 homo sapien
10	570	29.2	328	Q8BJR6	Q8BJR6 mus musculus
11	567	29.0	389	Q9PVX7	Q9PVX7 xenopus lae
12	561	28.7	320	Q7T0X2	Q7T0X2 xenopus lae
13	560.5	28.7	297	Q88781	Q88781 rattus ratt
14	558	28.6	317	Q9DGR3	Q9DGR3 xenopus lae
15	540	27.6	422	Q8WVCL	Q8WVCL homo sapien
16	540	27.6	581	Q9BYE2	Q9BYE2 homo sapien

ALIGNMENTS

RESULT 1

Q8NBY4 ID Q8NBY4 PRELIMINARY; PRT; 766 AA.
AC Q8NBY4;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hypothetical protein FLJ90661.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Placenta;
RA Isogai T., Ota T., Nishikawa T., Hayashi K., Otsuki T., Sugiyama T.,
RA Suzuki Y., Nagai K., Sugano S., Ishii S., Kawai-Hio Y., Saito K.,
RA Yamamoto J., Wakamatsu A., Nakamura Y., Kojima S., Nagahari K.,
RA Masubo Y., Ono T., Okano K., Yoshikawa Y., Aotsuka S., Sasaki N.,
RA Hattori A., Okumura K., Iwayanagi T., Ninomiya K.;
RA "NEDO human cDNA sequencing project";
RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1.
DR EMBL; AK075142; BAC11431.1; -
DR GO; GO:0004263; F:chymotrypsin activity; IEA.
DR GO; GO:0008233; F:peptidase activity; IEA.
DR GO; GO:0004295; F:trypsin activity; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro; IPR009003; Cys Ser trypsin.
DR InterPro; IPR001254; Peptidase S1.
DR InterPro; IPR001314; Peptidase_S1A.
DR Pfam; PF00089; trypsin; 3.
DR PRINTS; PR00722; CHYMOTRYPSIN.
DR SMART; SM00020; Tryp_SPC; 3.
DR PROSITE; PS00240; TRYPSIN_DOM; 3.
DR PROSITE; PS00135; TRYPSIN_SER; 1.
KW Hypothetical protein; Hydrolase; Protease; Serine protease.
SQ SEQUENCE 766 AA; 82564 MW; 3630D550CB06BD55 CRC64;

17 538.5 27.6 307 11 Q7TML0
18 538.5 27.6 558 4 Q86YM4
19 537 27.5 328 11 Q80Z40
20 537 27.5 537 4 Q9BYE1
21 530.5 27.2 311 11 Q80XZ3
22 530 27.1 855 11 Q9JUI7
23 529 27.1 336 11 Q80YD8
24 521 26.7 471 11 Q8CFE0
25 516 26.4 310 11 Q91XC4
26 510.5 26.1 572 11 Q8BIK6
27 510 26.1 310 11 Q9QY29
28 501 25.7 1059 4 Q7Z411
29 497.5 25.5 273 6 Q9XSM1
30 495 25.3 453 11 Q812A6
31 493.5 25.3 282 11 Q9D4I3
32 493.5 25.3 322 11 Q920S2
33 493 25.2 624 11 Q9DAT3
34 491 25.1 855 4 Q7Z410
35 489 25.0 275 4 Q86TM8
36 489 25.0 276 4 Q86UA5
37 487 24.9 624 11 Q91Y47
38 483.5 24.8 806 6 Q18783
39 479 24.5 329 13 Q42272
40 474.5 24.3 335 11 Q8VIF2
41 473.5 24.2 275 4 Q96RZ6
42 473 24.2 421 11 Q60491
43 472.5 24.2 624 6 Q95ME7
44 470.5 24.1 643 6 Q97506
45 469 24.0 638 11 Q8ROP5

Q7Tml0 mus musculus
Q86ym4 homo sapien
Q80z40 rattus norv
Q9byel homo sapien
Q80xz3 rattus norv
Q9jui7 rattus norv
Q80yd8 mus musculus
Q8cfe0 mus musculus
Q91xc4 mus musculus
Q8bik6 mus musculus
Q9qyz9 mus musculus
Q7z411 homo sapien
Q9xsm1 ovis aries
Q812a6 mus musculus
Q9d4i3 mus musculus
Q920s2 mus musculus
Q9dat3 mus musculus
Q7z410 homo sapien
Q86tm8 homo sapien
Q86ua5 homo sapien
Q91y47 mus musculus
Q18783 macropus eu
Q42272 xenopus lae
Q8vif2 mus musculus
Q96rz6 homo sapien
Q60491 cavia porce
Q95me7 oryctolagus
Q97506 sus scrofa
Q8rop5 mus musculus

Query Match 61.5%; Score 1202; DB 4; Length 766;
 Best Local Similarity 100.0%; Pred. No. 2.1e-88;
 Matches 220; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 90 MTNGLEPAEWSVLGVHSGDGLDGAHTRAVAAIVVPANYSQVELGADLALLRLASPA 149
 DB 1 MTNGLEPAEWSVLGVHSGDGLDGAHTRAVAAIVVPANYSQVELGADLALLRLASPA 60
 QY 150 SLGPVAVPVCPLPRASHRFVHGHTACWATGWDVQVQADPLPLPWLQVEVLRLIGATCQCL 209
 DB 61 SLGPVAVPVCPLPRASHRFVHGHTACWATGWDVQVQADPLPLPWLQVEVLRLIGATCQCL 120
 QY 210 YSQGPENLTQLPQMLCAGYFEGRTDCCQDSGGLPVCBEGGRWFQAGITSFQGGQR 269
 DB 121 YSQGPENLTQLPQMLCAGYFEGRTDCCQDSGGLPVCBEGGRWFQAGITSFQGGQR 180
 QY 270 RNRPGVTAATAYAWIREQVMGSRPGPAFTPOKQTQSD 309
 DB 181 RNRPGVTAATAYAWIREQVMGSRPGPAFTPOKQTQSD 220

RESULT 2

Q99L44 PRELIMINARY; PRT; 339 AA.
 AC Q99L44;
 DT 01-JUN-2001 (TremBLrel. 17, Created)
 DT 01-JUN-2001 (TremBLrel. 17, Last sequence update)
 DT 01-OCT-2003 (TremBLrel. 25, Last annotation update)
 DE Similar to protease, serine, 8 (Prostasin).
 GN PRS8.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Strausberg R.;
 RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=129/SvJ, and Swiss; TISSUE=Lung;
 RA Verghese G.M., Caughey G.H.;
 RT "Molecular cloning and characterization of mouse prostasin, a type I
 membrane-associated serine protease of the gamma-trypsin/prostasin
 gene family."
 RT Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RA Kitamura K., Takefumi N., Kimio T.;
 RT "mouse serine protease."
 RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
 CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1.
 DR EMBL; BC003851; AA003851.1; -.
 DR EMBL; AF378086; AAL06320.1; -.
 DR EMBL; AF378085; AAL06319.1; -.
 DR EMBL; AB038244; BAB82496.1; -.
 DR HSSP; P00734; IUVS.
 DR MGD; MGI:1923810; Prss8.
 DR GO; GO:0004263; F:chymotrypsin activity; IEA.
 DR GO; GO:0008233; F:peptidase activity; IEA.
 DR GO; GO:0004295; F:trypsin activity; IEA.
 DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
 DR InterPro; IPR009003; Cys_Ser_trypsin.
 DR InterPro; IPR001254; Peptidase_S1.
 DR InterPro; IPR001314; Peptidase_S1A.
 DR Pfam; PF00089; trypsin; 1.
 DR PRINTS; PR00722; CHYMOTRYPSIN.
 DR SMART; SM00020; Tryp_SPC; 1.
 DR PROSITE; PS00240; TRYPSIN_DOM; 1.
 DR PROSITE; PS00134; TRYPSIN_HIS; 1.
 DR PROSITE; PS00135; TRYPSIN_SER; 1.
 KW Hydrolase; Protease; Serine protease.
 SQ SEQUENCE 339 AA; 36216 MW; BC2DE88BFC057AF10 CRC64;

Query Match 32.3%; Score 630.5; DB 11; Length 339;
 Best Local Similarity 41.7%; Pred. No. 8.7e-43;
 Matches 145; Conservative 48; Mismatches 138; Indels 17; Gaps 9;

QY 1 MAQKGVLGPGQLGAVANSDSYSLYGLVPSG-PARG-PPYCGRP-EP SARIVGSGNAQPGT 57
 DB 1 MALRVGLGQLEAVT---ILLGLLQSGIRADGTASCGAVTQP--RITGGGSAKPGQ 55
 QY 58 WPMQVSLHHGGCH:CGSLIAPSVMVLSAAHCFMTNGTLEPAEWSVLGVHSGDGLDGA 117
 DB 56 WPMQVSLHHGGCH:CGSLIAPSVMVLSAAHCFMTNGTLEPAEWSVLGVHSGDGLDGA 112
 QY 118 HTRAVAAIVVPANYSQVELGADLALLRLASPASLGPVWPVCLPRASHRFVHGHTACWATG 177
 DB 113 VVHTVAQIITHSSVRESGSQDIALIRLSVPVTSRYIRPICLPAANASFFNGLHCTVTG 172
 QY 178 WGDVQVADPLPLPWLQVEVLRLIGATCQCLYQSGPFNLTQLPQMLCAGYFEGRTD 237
 DB 173 WGHVAPSVSLQTPRPLQOQLEVLISRETSCLYINAVPEEFTIQDMLCAGYVKGKGD 232
 QY 238 TCQDSGGPLVCBEGGRWFQAGITSFQGGRRNRPGVFTAVATYEAWEIRQVMGSEPGP 297
 DB 233 ACQDSGGPLSCPMEGIWYLAGIVSWGDACGAPNRPVYTLTSTYASWIIHHV--AEIQP 290
 QY 298 APPTQPKTQSD---CLHQTAFLDAR-ILLRPLSHISGVSTGKSL 341
 DB 291 RVPQTQESQDGLCHNHPVFSSAAAPKLLRPVLPFLGLTGLLSL 338

RESULT 3

Q8BUJ6 PRELIMINARY; PRT; 340 AA.
 AC Q8BUJ6;
 DT 01-MAR-2003 (TremBLrel. 23, Created)
 DT 01-MAR-2003 (TremBLrel. 23, Last sequence update)
 DT 01-OCT-2003 (TremBLrel. 25, Last annotation update)
 DE Prostasin.
 GN PRS8.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=Eye;
 RX MEDLINE=22354683; PubMed=12466851;
 RA The FANTOM Consortium,
 RA the RIKEN Genome Exploration Research Group Phase I & II Team;
 RT "Analysis of the mouse transcriptome based on functional annotation of
 60,770 full-length cDNAs."
 RL Nature 420:563-573 (2002).
 DR EMBL; AK078696; BAC37362.1; -.
 DR MGD; MGI:1923810; Prss8.
 DR GO; GO:0004263; F:chymotrypsin activity; IEA.
 DR GO; GO:0004295; F:trypsin activity; IEA.
 DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
 DR InterPro; IPR009003; Cys_Ser_trypsin.
 DR InterPro; IPR001254; Peptidase_S1.
 DR InterPro; IPR001314; Peptidase_S1A.
 DR Pfam; PF00089; trypsin; 1.
 DR PRINTS; PR00722; CHYMOTRYPSIN.
 DR SMART; SM00020; Tryp_SPC; 1.
 DR PROSITE; PS00240; TRYPSIN_DOM; 1.
 DR PROSITE; PS00134; TRYPSIN_HIS; 1.
 DR PROSITE; PS00135; TRYPSIN_SER; 1.
 SQ SEQUENCE 340 AA; 36503 MW; 3C2540E9B3A81C8A CRC64;

Query Match 31.4%; Score 613; DB 11; Length 340;

Best Local Similarity 41.5%; Pred. No. 2.2e-41;
 Matches 145; Conservative 46; Mismatches 140; Indels 18; Gaps 10;

QY 1 MAQKGVLGPGQLGAVANSDSYSLYGLVPSG-PARG-PPYCGRP-EP SARIVGSGNAQ-PG 56

```
Db 1 MALRVGLGLQLEAVT---ILLGLLQSGIRADGTASCGAVIQP-RITGGGQWQSG 55
QY 57 TWPQVSLHHGGHICGGSLIAPSWLAAHCFWMTGTLPEAAEWVLLGVHSDQGLDG 116
Db 56 QMPWQVSIYDGNHVGCGSLVSNKVVSAHCFPRHSRE---AYEVKLGDLHQLDSYND 112
QY 117 AHTRAAAIIVPANYSQVELGADLALLRLASPASLGPFAVMPVCLPRASHRFVHGTACWAT 176
Db 113 TVVHTVAQIITHSSYREESQGLALIRLSSPFFSYRIPICLPANASFPNGLHCTVT 172
QY 177 GMDVQVADPLPLPWVQLQVEVELLGEATCCQLYSQPGPNLTQLIPLGMLCAGYPEGRR 236
Db 173 GMDVAPSVSLQPRPLQLEVLLEISRETCSCLYNNAVPEPHITQQDMCLCAGYVKGK 232
QY 237 DTCQSGGGLVCEEGRWFOAGITSGFCGGRNRPVFTAVATYEAWEATREQVMSSEPG 296
Db 233 DACQSGGGLSLCPMGIMGYLAGVSWGDACGAPNRPVTVTLTSTYASVHHHV---AEIQ 290
QY 297 PAFPTQPKTQSD---CLHQTAFEDSAR-ILLRPLSHISVGVSTGKSL 341
Db 291 PRVVPQTQESQPDGHLNCHHPFSSAAAPKLLRPVLFPLGLGLLGLSL 339
RESULT 4
Q80WM7 PRELIMINARY; PRT; 277 AA.
AC Q80WM7;
DT 01-JUN-2003 (TREMBlrel. 24, Created)
DT 01-JUN-2003 (TREMBlrel. 24, Last sequence update)
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE Tryptase-6.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BALB/c; TISSUE=Testis;
RA Wong G.W., Yasuda S., Li L., Stevens R.L.;
RT "Cloning and characterization of mouse tryptase-6 (mt6).";
RL Submitted (MAR-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY262880; AAF20885.1; -.
DR GO; GO:0004263; F:chymotrypsin activity; IEA.
DR GO; GO:0004295; F:trypsin activity; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro; IPR009003; Cys Ser trypsin.
DR InterPro; IPR001254; Peptidase_S1.
DR InterPro; IPR001314; Peptidase_S1A.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PR00722; CHYMOTRYPSIN.
DR SMART; SM00020; Tryp_SPC; 1.
DR PROSITE; PS00240; TRYPSIN_DOM; 1.
DR PROSITE; PS00134; TRYPSIN_HIS; 1.
DR PROSITE; PS00135; TRYPSIN_SER; 1.
SQ SEQUENCE 277 AA; 29887 MW; 3310C067573E5E74 CRC64;
Query Match 30.9%; Score 603; DB 11; Length 277;
Best Local Similarity 46.4%; Pred. No. 1.1e-40;
Matches 122; Conservative 30; Mismatches 89; Indels 22; Gaps 6;
QY 38 CGRPEPSARIVGSSNAQPGWFWQVSLHHGGHICGGSLIAPSWLAAHCFWMTGTLPE 97
Db 25 CGQPRMSSRIVGGRDAQDGWFWQVSLHHGGHICGGSLIAPSWLAAHCFWMTGTLPE 77
QY 98 ----AAEWVLLGVHSDQGLDGADHTRAAAIIVPANYSQVELGADLALLRLASPASL 151
Db 78 RRVMPSEYVLLGALSLD--VRSHELLVPLVRLPPDYSEDEARGDLALLQLRFPVSL 135
QY 152 GPVAVMPVCLPRASHRFVHGTACWATGVDVQVADPLPLPWVQLQVEVELLGEATCCQLY- 210
Db 136 STRIQVCLPAGSHPPPGFCWWTGWSLSPGVPLPKRPGQVVRVPLLDLRACDLVH 195
```

```
QY 211 ---SOPGPENLTQLIPLGMLCAGYPEGREDTCCQSDSGGPLVCEEGRWFOAGITSGFGSC 267
Db 196 VGANVPOGRI---VLPGLCAGYRGHDKACQSGSGGLTCTWESGHVVLGVVSWGKGC 252
QY 268 GRNRPGVFTAVATYEAWEATREQV 290
Db 253 ALPNRPVTVNVAKYSPWIAQL 275
RESULT 5
Q80X17 PRELIMINARY; PRT; 331 AA.
AC Q80X17;
DT 01-JUN-2003 (TREMBlrel. 24, Created)
DT 01-JUN-2003 (TREMBlrel. 24, Last sequence update)
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE Tryptase 5.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BALB/c; TISSUE=Brain;
RA Wong G.W., Yasuda S., Stevens R.L., Li L.;
RT "Cloning and characterization of mouse tryptase-5 (mt5).";
RL Submitted (MAR-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY266139; AAP23216.1; -.
DR GO; GO:0004263; F:chymotrypsin activity; IEA.
DR GO; GO:0004295; F:trypsin activity; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro; IPR009003; Cys Ser trypsin.
DR InterPro; IPR001254; Peptidase_S1.
DR InterPro; IPR001314; Peptidase_S1A.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PR00722; CHYMOTRYPSIN.
DR SMART; SM00020; Tryp_SPC; 1.
DR PROSITE; PS00240; TRYPSIN_DOM; 1.
DR PROSITE; PS00134; TRYPSIN_HIS; 1.
DR PROSITE; PS00135; TRYPSIN_SER; 1.
SQ SEQUENCE 331 AA; 35640 MW; B1B36F32EF720377 CRC64;
Query Match 30.9%; Score 603; DB 11; Length 331;
Best Local Similarity 43.9%; Pred. No. 1.4e-40;
Matches 129; Conservative 38; Mismatches 115; Indels 12; Gaps 6;
QY 38 CGRPEPSARIVGSSNAQPGWFWQVSLHHGGHICGGSLIAPSWLAAHCFWMTGTLPE 97
Db 45 CGRPTSGRIVSGQDAQLGRWFWQVSVRENGARVCGSLIAEDWVLTAAHCFNQGS-- 102
QY 98 AAWSVLLG-VHSQDGLDGAHTRAAAIIVPANYSQVE-LGADLALLRLASPASLGPV 155
Db 103 -SIYTVLLGTISSEYEDNEPKELRAVAQIKPSYSADEHSGDIALVQLASISFNIDY 151
QY 156 WPVCLPRASHRFVHGTACWATGVDVQVADPLPLPWVQLQVEVELLGEATCCQLYSQ--- 212
Db 162 LPVCLPKPGDPLDPGTCWVTCWGHIGTNQPLPPPTLQELQVELIDAETCTNYQENSI 221
QY 213 PGFNLTLIPLGMLCAGYPEGREDTCCQSDSGGPLVCEEGRWFOAGITSGFGCGRRNR 272
Db 222 PGTEFY---ILRGLCAGFQEGKDACNGDSGGLVDCINDVMIQAGVYVSWGSDCALFKR 278
QY 273 PGVFTAVATYEAWEATREQVMSSEPGFAFPPTQPKTQSDCLHQTAFEDSARILLRP 326
Db 279 PGYTNVSVIISWI-QNTMWNLPMEGRGFPSPSLSGTPLLGLTSLSSAFLILGP 331
RESULT 6
Q8RIAG PRELIMINARY; PRT; 331 AA.
AC Q8RIAG;
DT 01-JUN-2002 (TREMBlrel. 21, Created)
DT 01-JUN-2002 (TREMBlrel. 21, Last sequence update)
```



```
DR PROSITE; PS00134; TRYPSIN_HIS; 1.
DR PROSITE; PS00135; TRYPSIN_SER; 1.
KW Hydrolase; Protease; Serine protease.
FT NON_TER 1
SQ SEQUENCE 327 AA; 35106 MW; 60458ED817AC1CF3 CRC64;

Query Match 30.2%; Score 590; DB 4; Length 327;
Best Local Similarity 46.3%; Pred. No. 1.5e-39;
Matches 119; Conservative 29; Mismatches 99; Indels 10; Gaps 3;

QY 38 CGRPEPARIIVCGSNAQPGTWPMQVSLHGGGHICGSLIAPSWLSAAHCFMTNLTLP 97
Db 75 CGPQRRSSRIIVGDRGDEWPMQASIQHGAHVCGSLIAPQWLVTAACHCFRRAL--- 131
QY 98 AAEWSVLLGVHSGDGLDGAHTRAVAAIVVPANYSQVELGADIALRLRLASPLGPAVWP 157
Db 132 PAEYRVLGALRLGSTSPRTLSVPVRVLLPPDYSEDGARGDLALLQLRRPPLSARVQ 191
QY 158 VCLPRASHRFVHGTACWATGWQVQADPLPLPWVLOEVELRLLGATCCQLY---SOP 213
Db 192 VCLPVPGARPPPGTPCRVTGWSGLRPGVLEFWRPLQGVVRVPLDLSRTCDGLYHVGADVP 251
QY 214 GPNLTLQILPGMLCAGYEGRRDTCQDGGGGLVCEEGRWFFQAGITSFEGCGRRNP 273
Db 252 QABRI---VLPGLCAGYQGHKDACQDGGGGLTCLQSGSWVLVGVSWGKGCALPNRP 308
QY 274 GVFTAVATYEAWIREQV 290
Db 309 GYVTSVATYSPWLOARV 325

RESULT 9
ID Q96RZ8 PRELIMINARY; PRT; 321 AA.
AC Q96RZ8;
DT 01-DEC-2001 (TREMELrel. 19, Created)
DT 01-DEC-2001 (TREMELrel. 19, Last sequence update)
DT 01-OCT-2003 (TREMELrel. 25, Last annotation update)
DE HS transmembrane tryptase, gene name TMT, AF175522_1.
GN TRPA.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21096910; PubMed=11157797;
RA Daniels R.J., Peden J.F., Lloyd C., Horsley S.W., Clark K.,
RA Tufarelli C., Kearney L., Buckle V.J., Doggett N.A., Flint J.,
RA Higgs D.R.;
RT "sequence, structure and pathology of the fully annotated terminal 2
RT Mb of the short arm of human chromosome 16.";
RL Hum. Mol. Genet. 10:339-352(2001).
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1.
DR EMBL; AE006466; AK61269.1; -.
DR HSSP; P00761; 1ANI.
DR GO; GO:0016021; C:integral to membrane; IEA.
DR GO; GO:0004263; F:chymotrypsin activity; IEA.
DR GO; GO:0008233; F:peptidase activity; IEA.
DR GO; GO:0004295; F:trypsin activity; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro; IPR009003; Cys Ser trypsin.
DR InterPro; IPR001254; Peptidase_S1.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PR00722; CHYMOTRYPSIN.
DR SMART; SM0020; Tryp_SPC; 1.
DR PROSITE; PSS0240; TRYPSIN_DOM; 1.
DR PROSITE; PS00134; TRYPSIN_HIS; 1.
KW Hydrolase; Protease; Serine protease; Transmembrane.
SQ SEQUENCE 321 AA; 33829 MW; FFF5089EDC4FC73D CRC64;

Query Match 29.8%; Score 582; DB 4; Length 321;

Best Local Similarity 46.0%; Pred. No. 6.6e-39;
Matches 128; Conservative 35; Mismatches 87; Indels 28; Gaps 11;

QY 36 PYCGRPEPS---ARIVGGSSNAQPGTWPMQVSLHGGGHICGSLIAPSWLSAAHCFMTN 92
Db 24 FCGRQVSDAGGRIVGHAAPAGAWPMQASLRLRRVHVCSSLLSPQWLVTAACHCF--S 81
QY 93 GTLEPAEWSVLLGVHSGDGLD---GAHTRAVAAIVVPANYS-QVELGADIALRLRLASP 148
Db 82 GSLN-SSDYQVHL-----GELEITLSPHFSTVRQIILHSSPSGQGTSGDIALVELSV 134
QY 149 ASLGPAAVFPVCLPRASHRFVHGTACWATGWQVQADPLPLPWVLOEVELRLLGATCC 208
Db 135 VTLSSRIILPVCLPEASDDPCFCIRCWVTGWYTRREGELPPPYSLREVKVSVVDVTCR 194
QY 209 LYSQPGPNLTLQILPGMLCAGYEGRRDTCQDGGGGLVCEEGRWFFQAGITSFEGCG 268
Db 195 DY--FQPGSILQ--PDMICA---RPGDACQDGGGGLVQVGNWVQAGIVSWGEGCG 247
QY 269 RNRPGVFTAVATYEAWIREQVM---GSEPGPAFTQP 303
Db 248 RNRPGVTRVPAVYVNWIRRHITASGSESG--YPLRP 283

RESULT 10
ID Q8BJR6 PRELIMINARY; PRT; 328 AA.
AC Q8BJR6;
DT 01-MAR-2003 (TREMELrel. 23, Created)
DT 01-MAR-2003 (TREMELrel. 23, Last sequence update)
DT 01-OCT-2003 (TREMELrel. 25, Last annotation update)
DE Similar to MARAPASIN precursor (Channel-activating protease 2-like
DE protein) (Pancreasin).
GN MPN.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Thymus;
RX MEDLINE=22354883; PubMed=12466851;
RA The FANTOM Consortium,
RA the RIKEN Genome Exploration Research Group Phase I & II Team;
RT "Analysis of the mouse transcriptome based on functional annotation of
RT 60,770 full-length cDNAs.";
RL Nature 420:563-573(2002).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=12986/SvevTac;
RX MEDLINE=22439795; PubMed=12441343;
RA Bhagwandin V.J., Hau L.W., Mallen-St Clair J., Wolters P.J.,
RA Caughey G.H.;
RT "Structure and activity of human pancreasin, a novel tryptic serine
RT peptidase expressed primarily by the pancreas.";
RL J. Biol. Chem. 278:3363-3371(2003).
DR EMBL; AK080281; BAC37864.1; -.
DR EMBL; AY162410; AAC17162.1; -.
DR EMBL; AF542056; AAC27572.1; -.
DR MGD; MGI:2450123; Mpn.
DR GO; GO:0004263; F:chymotrypsin activity; IEA.
DR GO; GO:0008233; F:peptidase activity; IEA.
DR GO; GO:0004295; F:trypsin activity; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro; IPR009003; Cys Ser trypsin.
DR InterPro; IPR001254; Peptidase_S1.
DR InterPro; IPR001314; Peptidase_S1A.
```

```
DR Pfam: PF00089; trypsin; 1.
DR PRINTS: PR00722; CHYMOTRYPSIN.
DR SMART: SM00020; Tryp_SPC; 1.
DR PROSITE; PS00240; TRYPSIN_DOM; 1.
DR PROSITE; PS00134; TRYPSIN_HIS; 1.
DR PROSITE; PS00135; TRYPSIN_SER; 1.
DR PROTEASE.
SQ SEQUENCE 328 AA; 35789 MW; DCOB20F1AB3EB840 CRC64;

Query Match      29.2%; Score 570; DB 11; Length 328;
Best Local Similarity 41.3%; Pred. No. 6.2e-38;
Matches 131; Conservative 38; Mismatches 126; Indels 22; Gaps 8;

QY 26 LVPSGP--ARGPPYCGRPEPSARIIVGGSNAQGTWPQVSLHGGCHICGSLIAPSWVL 83
DB 15 LRSITEGARTLRACHPKFNMVGGENALEGEWPQVSIQRNGIHFCGSLIAPTWWL 74
QY 84 SAAHCFMTNGTLEPAAEVSLGVHSQDGLDGAHTRAVAAIVPAN--YSQVELGADLA 141
DB 75 TAAHCFSTSDI---SIYQVLLGALKLQOP--GPHALYVPVQVKSNPQYQGMASADVA 129
QY 142 LLRLASPASLPAVWPVCLPRASHRFVHCTACWATGWDVQADPLPLPWVLOEVELRL 201
DB 130 LVELOGPVTFNIIPLVCLPDSVIFESGMNCWVTGWSGSPSQDLRPNRVLQKLAAPLI 189
QY 202 GEATCQCLYSQPGFENLIQ--ILPGMLCAGYPEGRDTCQDGGGLVCEEGGRWFQAGI 260
DB 190 DTPKCNLLYNKDVESDFQKTKIDDMLCAGFAEGKDKCKDGGGLVCLVDSQVQAGV 249
QY 261 TSFGGCGRRNRPVFTAVTAYEAMIREQVMGSE--PGPAFFTPQPK-----TQSD 309
DB 250 ISWGGCGARRNPGVYIRVTSHHKWIHQIPELQFQGRAGTQCCQKDSQGOORLAGNSAP 309
QY 310 CLHQTAFL-LDSARILLR 325
DB 310 CLAAHAWLALGALLR 326

RESULT 11
Q9PVX7 PRELIMINARY; PRT; 389 AA.
AC Q9PVX7;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Epidermis specific serine protease.
GN XEPSIN.
OS Xenopus laevis (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae;
OC Xenopodinae; Xenopus.
OX NCBI_TaxID=8355;
RN [1]
RP SEQUENCE FROM N.A.
RA Yamada K.;
RT "The expression control of xepsin by non-axial and planar
RT posteriorizing signals in Xenopus epidermis."
RL Submitted (OCT-1998) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1.
DB EMBL; AB018694; BAA84941.1; -
DB HSSP; P00763; IDPO.
DR GO; GO:0004263; F:chymotrypsin activity; IEA.
DR GO; GO:0008233; F:peptidase activity; IEA.
DR GO; GO:0044295; F:trypsin activity; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro; IPR009003; Cys Ser trypsin.
DR InterPro; IPR01254; Peptidase_S1.
DR InterPro; IPR001314; Peptidase_S1A.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PR00722; CHYMOTRYPSIN.
DR SMART; SM00020; Tryp_SPC; 1.
DR PROSITE; PS00240; TRYPSIN_DOM; 1.
DR PROSITE; PS00134; TRYPSIN_HIS; 1.

DR PROSITE; PS00135; TRYPSIN_SER; 1.
DR HYDROLASE; Protease; Serine protease.
SQ SEQUENCE 389 AA; 42375 MW; B31F8A2F5D1F6E3 CRC64;

Query Match      29.0%; Score 567; DB 13; Length 389;
Best Local Similarity 40.2%; Pred. No. 1.3e-37;
Matches 109; Conservative 45; Mismatches 97; Indels 20; Gaps 5;

QY 38 CGRPEPSARIIVGGSNAQGTWPQVSLHGGCHICGSLIAPSWVLSAAHCFMTNGTLEP 97
DB 17 CGVPVSNRIIVGGMDSKRGWFWQVLSYKSDSICGSLITDSWMTAAHCL---DSLQ- 72
QY 98 AAWSVLGVHSQDGLDGAHTRAVAAIVPANYSQVELGADLALLRLASPASLPAVWP 157
DB 73 VSYITVYLGAQLSAPDNSTVSRGKSIKHPDFQEGSGDIALIELEKPTPTPILP 132
QY 158 VCLPRASHRFVHCTACWATGWDVQADPLPLPWVLOEVELRLGEATCQCLYSQPGFN 217
DB 133 ICLPSQDVFQAGTMCWVTGWNIOEGTFLPSPTKIOAEVAIIDSSVCGTMYES----- 187
QY 218 LTLQILP-----GMLCAGYPEGRDTCQDGGGLVCEEGGRWFQAGITSFGGCGRR 270
DB 198 -SLGYIPDFEIQEDMVCAGYKEGIDACQDGGGLVGVNWNVWLQGLGVSWGVCRAEP 246
QY 271 NRPQVFTAVTAYEAMIREQV---MGSEPGPA 298
DB 247 NRPQVYTKVQYQDWLKNVPLIVFSEEGPS 277

RESULT 12
Q7TOX2 PRELIMINARY; PRT; 320 AA.
AC Q7TOX2;
DT 01-OCT-2003 (TrEMBLrel. 25, Created)
DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hypothetical protein.
OS Xenopus laevis (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae;
OC Xenopodinae; Xenopus.
OX NCBI_TaxID=8355;
RN [1]
RP SEQUENCE FROM N.A.
RA Klein S.L., Strausberg R.L., Wagner L., Pontius J., Clifton S.W.,
RA Richardson P.;
RT "Genetic and genomic tools for Xenopus research: The NIH Xenopus
RT initiative."
RL Dev. Dyn. 225:384-391 (2002).
RN [2]
RP SEQUENCE FROM N.A.
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Strausberg R.L., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Klausner R.D., Collins F.S., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Wang J., Wang J., Hsieh P.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hong L.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Udwin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullany S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettman M., Madan A.C., Rodrigues S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA Krzywinski M.I., Skalska U., Smailus D.E., Scherch A., Schein J.E.,
RA Jones S.J., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
```


Query Match 28.6%; Score 558; DB 13; Length 317;
Best Local Similarity 39.2%; Pred. No. 5.6e-37;
Matches 116; Conservative 45; Mismatches 119; Indels 16; Gaps 5;

QY 35 PPYGRPPSARIIVGGSSNAQPGTWQVLSLHGGCHICGSLIAPSWLVAHCF---M 90
DB 29 PPLCGSPFSSRIIVGGTITROGANFQWVSLFNGSHICGSLISDQWLLTATHCIEHPDL 88
QY 91 TNGTLEPAAEWSVLGVHSQDGLDGAHTRAVAIVPANYQVELGADLALLRIASPAS 150
DB 89 PSGC-----GVLGAYQLVKNPHEMTVKVDIIYINSEFNGPTSGDIALLLKLSPIK 141
QY 151 LGPAWVPVCLPRASHRFVHGTACWATGWDVQVQADPLPLWVLOVELRLLGATCCCLY 210
DB 142 FTEVILPCLPASPTVFFSGTECMITGQGTGSEVPLQYPATLQKVMVPIINRSCENY 201
QY 211 SQPGFNLIT-LQILPGMLCAGYPEGRRDTCQDGGPLVCEBGRWFOAGITSEFGGGR 269
DB 202 HINSVISETEILIQSDQICAGYQAGQKQCGQDGGPLVCKIQGFYQAGIVSWGERCAA 261
QY 270 RNRPGVFTAVATYEAMIREQ-VMGSEPGAPPTQPKTQSCDLHOTAFDLSARILL 324
DB 262 KNRPGVTFVPAVETWISERSVISFKP---FTSSSPSSSSVLRASALLGVSLIL 314

RESULT 15

Q8WVC1 PRELIMINARY; PRT; 422 AA.

AC Q8WVC1;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hypothetical protein (Fragment).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Colon;
RA Strausberg R.;
RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1.
DR EMBL; BC018146; AAH18146.1; -
DR HSSP; P00761; IANI.
DR GO; GO:0004263; F:chymotrypsin activity; IEA.
DR GO; GO:0008233; F:peptidase activity; IEA.
DR GO; GO:0004295; F:trypsin activity; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro; IPR009003; Cys_Ser_trypsin.
DR InterPro; IPR002172; LDL_receptor_A.
DR InterPro; IPR001254; Peptidase_S1.
DR InterPro; IPR001314; Peptidase_S1A.
DR Pfam; PF00057; ldl_recept_a; 4.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PR00722; CHYMOTRYPSIN.
DR PRINTS; PR00261; LDLRECEPTOR.
DR SMART; SM00192; LDLA; 4.
DR SMART; SM00020; Tryp_SPC; 1.
DR PROSITE; PS01209; LDLRA_1; 2.
DR PROSITE; PS50068; LDLRA_2; 4.
DR PROSITE; PS50240; TRYPSIN_DOM; 1.
DR PROSITE; PS00134; TRYPSIN_HIS; 1.
DR PROSITE; PS00135; TRYPSIN_SER; 1.
KW Hypothetical protein; Hydrolase; Protease; Serine protease.
FT NON_TER
SQ SEQUENCE 422 AA; 46257 MW; 2C99875D1B58B319 CRC64;

Query Match 27.6%; Score 540; DB 4; Length 422;
Best Local Similarity 40.2%; Pred. No. 2.1e-35;
Matches 113; Conservative 39; Mismatches 95; Indels 34; Gaps 7;

QY 32 ARGPYCGRDEP-----SARIVGGSSNAQPGTWQVLSLHH-GGGHI 71
DB 148 SKGNPECDKEDSCDSDKDCDGLRSFTRQARVVGTTADDEGENWQVSLHAGGGHI 207
QY 72 CGSLIAPSWLVAHCFMTNG--TLEPAAEWSVLGVHSQ-DGPLDGAHTRAVAIVP 128
DB 208 CGASLISPNNLVSAAHCIIDRGFRYSDDPTQWTAFLGLHDQSSAPGVQERLKRILSH 267
QY 129 ANYSQVELGADLALLRIASPASLGPVWPVCLPRASHRFVHGTACWATGWDVQVQADPLP 188
DB 268 PTFNDFDYDIALLELEKPAEYSSWVRPICLPDASHVFFAGKAIWYTGHTQYGGTGA 327
QY 189 LPWVLOVELRLLGATCCCLYQSGPGFNLITLQILPGMLCAGYPEGRRDTCQDGGPL- 247
DB 328 L--ILQKGEIRVINQTTCCENLLPQ-----QITPRMCMVGLSGGVDSQCGDSGGPLS 377
QY 248 VCEBGRWFOAGITSEFGGGRNRPGVFTAVATYEAMIRE 288
DB 378 SVEADGRIFQAGVYVWGDGCAQRNKGVTYTRLPFLFRDWIKE 418

Search completed: February 25, 2004, 15:49:33

Job time : 87 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 25, 2004, 15:45:03 ; Search time 44 Seconds
(without alignments)
418.874 Million cell updates/sec

Title: US-10-037-417-46

Perfect score: 1953
Sequence: 1 MAQGVILGFCQLGAVANSDS.....TKSLVLPWLSPLSLGLGWF 357

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents AA.*
1: /cgn2_6/ptodata/2/iaa/5A COMB.pep.*
2: /cgn2_6/ptodata/2/iaa/5B COMB.pep.*
3: /cgn2_6/ptodata/2/iaa/6A COMB.pep.*
4: /cgn2_6/ptodata/2/iaa/6B COMB.pep.*
5: /cgn2_6/ptodata/2/iaa/PCTUS COMB.pep.*
6: /cgn2_6/ptodata/2/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	602	30.8	299	3	US-08-944-483-66
2	590	30.2	284	4	US-09-387-375-7
3	563	28.8	290	4	US-09-386-653A-7
4	557	28.5	316	4	US-09-387-375-9
5	550.5	28.2	319	4	US-09-386-642-12
6	550	28.2	328	4	US-09-386-642-11
7	544	27.9	317	4	US-09-386-629-7
8	544	27.9	317	4	US-09-907-794A-263
9	544	27.9	317	4	US-09-905-125A-263
10	544	27.9	317	4	US-09-902-775A-263
11	540	27.6	855	2	US-09-027-337-2
12	540	27.6	855	4	US-09-644-600-2
13	540	27.6	855	4	US-09-654-600A-2
14	534.5	27.4	314	3	US-09-008-271A-3
15	534.5	27.4	314	4	US-09-907-794A-257
16	534.5	27.4	314	4	US-09-905-125A-257
17	534.5	27.4	314	4	US-09-902-775A-257
18	533.5	27.3	314	4	US-09-023-942A-6
19	531	27.2	315	4	US-09-386-653A-9
20	529	27.1	285	4	US-09-023-942A-26
21	526.5	27.0	312	4	US-09-023-942A-4
22	506.5	25.9	902	4	US-09-644-600-10
23	506.5	25.9	902	4	US-09-654-600A-10
24	502	25.7	327	4	US-09-386-629-8
25	499	25.6	276	2	US-09-016-366A-15
26	499	25.6	276	2	US-08-978-404B-21
27	495	25.3	454	3	US-09-518-046-2

28	491.5	25.2	274	2	US-08-978-404B-5	Sequence 5, Appli
29	491	25.1	235	3	US-08-944-483-65	Sequence 65, Appli
30	490.5	25.1	376	4	US-09-820-002-2	Sequence 2, Appli
31	490.5	25.1	416	2	US-09-000-846-2	Sequence 2, Appli
32	490.5	25.1	417	4	US-09-820-002-4	Sequence 4, Appli
33	487.5	25.0	249	3	US-09-079-970A-5	Sequence 5, Appli
34	487	24.9	306	4	US-09-386-642-53	Sequence 53, Appli
35	485.5	24.9	273	2	US-09-016-366A-19	Sequence 19, Appli
36	485.5	24.9	273	2	US-08-978-404B-14	Sequence 14, Appli
37	484.5	24.8	274	2	US-09-016-366A-21	Sequence 21, Appli
38	484.5	24.8	274	2	US-08-978-404B-16	Sequence 16, Appli
39	482.5	24.7	244	4	US-09-601-318-4	Sequence 4, Appli
40	482.5	24.7	244	4	US-09-601-318-5	Sequence 5, Appli
41	482.5	24.7	244	4	US-09-601-318-6	Sequence 6, Appli
42	482.5	24.7	244	4	US-09-601-318-7	Sequence 7, Appli
43	482.5	24.7	245	3	US-09-079-970A-6	Sequence 6, Appli
44	482.5	24.7	245	4	US-09-601-318-1	Sequence 1, Appli
45	479	24.5	256	2	US-09-027-337-3	Sequence 3, Appli

ALIGNMENTS

RESULT 1
US-08-944-483-66
; Sequence 66, Application US/08944483
; Patent No. 6232456

GENERAL INFORMATION:

APPLICANT: COHEN, MAURICE
APPLICANT: COLPITTS, TRACEY L.
APPLICANT: FRIEDMAN, PAULA N.
APPLICANT: GRANADOS, EDWARD N.
APPLICANT: KLASS, MICHAEL R.
APPLICANT: RUSSELL, JOHN C.
APPLICANT: STEWART, KENT D.
APPLICANT: STROUPE, STEVEN D.
TITLE OF INVENTION: NOVEL SERINE PROTEASE REAGENTS
TITLE OF INVENTION: AND METHODS USEFUL FOR DETECTING AND TREATING DISEASES
TITLE OF INVENTION: OF THE PROSTATE
NUMBER OF SEQUENCES: 76
CORRESPONDENCE ADDRESS:
ADDRESSEE: Abbott Laboratories
STREET: 100 Abbott Park Road
CITY: Abbott Park
STATE: IL
COUNTRY: USA

ZIP: 60064-3500
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/944,483
FILING DATE:
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Becker, Cheryl L.
REGISTRATION NUMBER: 35,441
REFERENCE/DOCKET NUMBER: 6183.US.01
TELECOMMUNICATION INFORMATION:
TELEPHONE: 847/935-1729
TELEFAX: 847/938-2623
TELEX:

INFORMATION FOR SEQ ID NO: 66:
SEQUENCE CHARACTERISTICS:
LENGTH: 299 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear

```

QY 274 GVFTAVATYEAATREQV 290
   ||:||||| ||: ||
Db 262 GVTTSVATYSPWIQARV 278

RESULT 3
US-09-386-653A-7
; Sequence 7, Application US/09386653A
; Patent No. 6485864
; GENERAL INFORMATION:
; APPLICANT: Andrade-Gordon, Patricia
; APPLICANT: Darrow, Andrew
; APPLICANT: Qi, Jian-shen
; TITLE OF INVENTION: DNA encoding the novel human serine
; TITLE OF INVENTION: Protease T
; FILE REFERENCE: ORT-1032
; CURRENT APPLICATION NUMBER: US/09/386,653A
; CURRENT FILING DATE: 1999-08-31
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 7
; LENGTH: 290
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-386-653A-7

Query Match 28.8%; Score 563; DB 4; Length 290;
Best Local Similarity 42.2%; Pred. No. 8.6e-44;
Matches 117; Conservative 32; Mismatches 114; Indels 14;

QY 32 ARGPPYCGRPESARIVGGNAQPGTWPMQVSLHHGGHICGSLIAPSWLSAAH
   |: ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| :
Db 20 AKAATACGRPRMLNRMVGGQDTQEGEWPQVSTIQRNGSHFCGSLIAEQWLTAARH
   |: ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| :
QY 92 NQTLPEAEMSVLGVSHSQDGLDGATRAVAALVVPAN--YSQVELGADLALLRL
   |: ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| :
Db 78 RNTSETSL-YQVLGARGQLVQP--GPHMYARVRQVESPLPQCTASSADVALVEL
   |: ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| :
QY 150 SLGPVAVPVCLPRASHRFVHGTACWATGWDQVQADPLPLPWVLQVEVRLILGEAT
   |: ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| :
Db 135 PFTNYILPVCLPDPSPVIFETGMNCWVTGSPSEEDLLPEPRILQKLAIVPIIDTFK
   |: ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| :
QY 210 YSQPGFFNLTQLIPG-MLCAGYEGEGRDTCQDGGGLVCEGGRWFOAGITTSF
   |: ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| :
Db 195 YSKDTEFGYQPTKINDMLCAGFEKGKDKACKGSGGLPVCLVGGQSLQAGVISM
   |: ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| :
QY 269 RNRRCGVFTAVATYEAATREQVMSSEFGPAFTPOK 305
   ||:||||| ||: ||||| : ||||| : ||||| : ||||| : ||||| :
Db 255 RQNRFGVIRVTAHNWIIHRII-----PKLQFQPAR 285

RESULT 4
US-09-387-375-9
; Sequence 9, Application US/09387375
; Patent No. 6485957
; GENERAL INFORMATION:
; APPLICANT: Darrow, Andrew
; APPLICANT: Andrade-Gordon, Patricia
; APPLICANT: Qi, Jensen
; TITLE OF INVENTION: DNA Encoding the Human Serine
; TITLE OF INVENTION: Protease EOS
; FILE REFERENCE: ORT-1031
; CURRENT APPLICATION NUMBER: US/09/387,375
; CURRENT FILING DATE: 1999-08-31
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 9
; LENGTH: 316
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Amino acid
; OTHER INFORMATION: sequence of EOS zymogen fusion gene

```

US-09-387-375-9

Query Match 28.5%; Score 557; DB 4; Length 316;
Best Local Similarity 45.8%; Pred. No. 3.5e-43;
Matches 114; Conservative 27; Mismatches 98; Indels 10; Gaps 3;

QY 46 RIVGSSNAQPGTWPQVSLHHGGHICGSLIAPSWLSAAHCFMTNGTLEPAAEWSVLL 105
DB 51 KIVGGYALEGQWPMQVSIYEGVHVCVGSIIAPQWVLTAAHCFPRAL---PAEYRVRL 107

QY 106 GVHSQDGLDGAHTRAVAAIVVPANYSQVELGADLALLRLGEATCQCLY----SQGPFNLTLQ 165
DB 108 GALLRGSTSPRTLSVPVRRVLLPPDYSEDGARGDLALLQLRRFVPLSARVQVCLVPFCA 167

QY 166 RVHGTACWATGWDVQADPLPLPWVLOEVELRLIGEATCQCLY----SQGPFNLTLQ 221
DB 168 RPPPGTFCRVTCGWSLPGVPLFEWRPLQGVVPLDSDTCGLYHVGADVPQARI--- 224

QY 222 ILPGMLCAGYPEGRRDTCQDGGPLVCEGGRWFQAGITSFGCGRRNRPGVFTAVAT 281
DB 225 VLPGLCAGYPOGHRKDACQDGGPLTCLQSGSWLVGVVSGKGCALPNRPGVITSVAT 284

QY 282 YEAWIREQV 290
DB 285 YSPWIOARV 293

RESULT 5

US-09-386-642-12
; Sequence 12, Application US/09386642
; Patent No. 6420157
; GENERAL INFORMATION:
; APPLICANT: Darrow, Andrew
; APPLICANT: Qi, Jensen
; APPLICANT: Andrade-Gordon, Patricia
; TITLE OF INVENTION: Zymogen Activation System
; FILE REFERENCE: ORT-1028
; CURRENT APPLICATION NUMBER: US/09/386,642
; CURRENT FILING DATE: 1999-08-31
; NUMBER OF SEQ ID NOS: 60
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 12
; LENGTH: 319
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Fusion gene
; OTHER INFORMATION: with homo sapien serine protease catalytic domain
US-09-386-642-12

Query Match 28.2%; Score 550.5; DB 4; Length 319;
Best Local Similarity 39.1%; Pred. No. 1.4e-42;
Matches 116; Conservative 42; Mismatches 116; Indels 23; Gaps 4;

QY 38 CGRP-----BPSARIVGSSNAQPGTWPQVSLHHGGHICGSLIAPSWV 82
DB 19 CGVPDYKDDDDAALAAFPDDDDKIVGGYALEAGQWPMQVSIYEGVHVCVGSIIAPSWV 78

QY 83 LSAHCFMTNGTLEPAAEWSVLLGVHSQDGLDGAHTRAVAAIVVPANYSQVELGADLAL 142
DB 79 LSAHCFPFSEHKE---AYEVKLGALQDLSYSDAKVSTLKDIIHPHSYLOEGSQGDIAL 135

QY 143 LRLASPSLGPVAVPVCIPRASHRFVHTCAWATGWDVQADPLPLPWVLOEVELRLIG 202
DB 136 LQLSRPTFSYIRIPICLPANAFSPNGLHCTVTGKHVAPSVSLTPKPLQQLLEVPLIS 195

QY 203 EATCQCLYSQPGFNLTLQILFGMLCAGYPEGRRDTCQDGGPLVCEGGRWFQAGITS 262
DB 196 RETCNCLYNIDAKPEEPHFVQEDMVCAGYVEGGKDACQDGGPLSCPVEGLWLTIGVS 255

QY 263 FFGCGRRNRPGVFTAVATYEAWEIREQVMGSEPGFAFTQPKTQSD---CLHQTAF 316
DB 256 WGDACGAERNRPGVFTAVATYEAWEIREQVMGSEPGFAFTQPKTQSD---CLHQTAF 310

RESULT 6

US-09-386-642-11
; Sequence 11, Application US/09386642
; Patent No. 6420157
; GENERAL INFORMATION:
; APPLICANT: Darrow, Andrew
; APPLICANT: Qi, Jensen
; APPLICANT: Andrade-Gordon, Patricia
; TITLE OF INVENTION: Zymogen Activation System
; FILE REFERENCE: ORT-1028
; CURRENT APPLICATION NUMBER: US/09/386,642
; CURRENT FILING DATE: 1999-08-31
; NUMBER OF SEQ ID NOS: 60
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 11
; LENGTH: 328
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Fusion gene
; OTHER INFORMATION: with homo sapien serine protease catalytic domain
US-09-386-642-11

Query Match 28.2%; Score 550; DB 4; Length 328;
Best Local Similarity 41.2%; Pred. No. 1.6e-42;
Matches 113; Conservative 41; Mismatches 112; Indels 8; Gaps 3;

QY 46 RIVGSSNAQPGTWPQVSLHHGGHICGSLIAPSWLSAAHCFMTNGTLEPAAEWSVLL 105
DB 51 KIVGGYALEAGQWPMQVSIYEGVHVCVGSIIAPSWVLTAAHCFPRAL---AYEVKL 107

QY 106 GVHSQDGLDGAHTRAVAAIVVPANYSQVELGADLALLRLGEATCQCLYSQPGFNLTLQ 165
DB 108 GAHQLDSYSDAKVSTLKDIIHPHSYLOEGSQGDIALQLSRPTFSYIRIPICLPANA 167

QY 166 RVHGTACWATGWDVQADPLPLPWVLOEVELRLIGEATCQCLYSQPGFNLTLQ 225
DB 168 SPFNGLHCTVTGKHVAPSVSLTPKPLQQLLEVPLISRETCLNIDAKPEEPHFVQED 227

QY 226 MLCAGYPEGRRDTCQDGGPLVCEGGRWFQAGITSFGCGRRNRPGVFTAVATYEA 285
DB 228 MVCAGYVEGGKDACQDGGPLSCPVEGLWLTIGVSWGDACGAERNRPGVFTAVATYEA 287

QY 286 IREQVMGSEPGFAFTQPKTQSD---CLHQTAF 316
DB 288 IQSKV--TELQPRVVPQTQESQPSDNLGSHLAF 319

RESULT 7

US-09-386-629-7
; Sequence 7, Application US/09386629
; Patent No. 6426199
; GENERAL INFORMATION:
; APPLICANT: Darrow, Andrew L.
; APPLICANT: Qi, Jensen
; APPLICANT: Andrade-Gordon, Patricia
; TITLE OF INVENTION: Identification and Characterization of the complementary
; TITLE OF INVENTION: DNA encoding the novel human serine protease C-E
; FILE REFERENCE: ORT-1030
; CURRENT APPLICATION NUMBER: US/09/386,629
; CURRENT FILING DATE: 1999-08-31
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 7
; LENGTH: 317
; TYPE: PRT
; ORGANISM: Homo sapiens
; OTHER INFORMATION: Hom
US-09-386-629-7

Query Match 27.9%; Score 544; DB 4; Length 317;
Best Local Similarity 39.3%; Pred. No. 5.4e-42;

Matches 114; Conservative 41; Mismatches 123; Indels 12; Gaps 4;
QY 7 LGPGQLGAVANSYSYLYGLVPSGARGPPYCGRPESARIVGGSNAQPGTWPQVSLHH 66
Db 10 LGGCGCLGTTSLLLASTAILNAARIIPVPACGKPOQLNRVVGEDSTDSEWPIVSIQK 69
QY 67 GGGHICGSLIAPSWVLSAAHCFMTNGTLEPAAEWSVLLGVHSDGPLDGAHTRAVAAIV 126
Db 70 NGTHCAGSLTSSRWITAAHCFKDN--LNKPYLFSVLLGAWQLGNFGRSQKVGVAWVE 127
QY 127 VPANYSQVE--LGADLALIRLASPASLGPAVMPVCLPRASHRFVHGTACWATGWDVQAD 185
Db 128 PHFYVSWKEGACADIALVRLERSIQFSERVLPICLPDASIHLPNTHCWSGSIQGV 187
QY 186 PLPLPWLQVVELRLGATCQCLY---SQGPNLTLQILPGMLCAGYPSGRDTCQGD 242
Db 188 PLPHPQTLQKLKVPIDSEVCSHLYWRGAGQGP-----ITEDMLCAGYLEGERDACLGD 241
QY 243 SGGPLVCEEGRGWFOAGITTSFGCGRNRPGVFTAVATYEAWIREQVMG 292
Db 242 SGGPLMCQVDGAWLLAGLIISWEGCAERNRFGVYISLSAHSRWSVEKIVQG 291

RESULT 8

US-09-907-794A-263
; Sequence 263, Application US/09907794A
; Patent No. 6635468
; GENERAL INFORMATION:
; APPLICANT: Genentech, Inc.
; APPLICANT: Ashkenazi, Avi
; APPLICANT: Botstein, David
; APPLICANT: Desnoyers, Luc
; APPLICANT: Eaton, Dan L.
; APPLICANT: Ferrara, Napoleone
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Fong, Sherman
; APPLICANT: Gerber, Hanspeter
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, A.
; APPLICANT: Godowski, Paul J.
; APPLICANT: Grimaldi, Christopher J.
; APPLICANT: Hillan, Kenneth, J.
; APPLICANT: Kljavin, Ivar J.
; APPLICANT: Mather, Jennie P.
; APPLICANT: Pan, James
; APPLICANT: Paoni, Nicholas F.
; APPLICANT: Roy, Margaret Ann
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Williams, P. Mickey
; APPLICANT: Wood, William, I.
; TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
; TITLE OF INVENTION: Acids Encoding the Same
; FILE REFERENCE: 10466-14
; CURRENT APPLICATION NUMBER: US/09/907,794A
; CURRENT FILING DATE: 2001-07-17
; PRIOR APPLICATION NUMBER: PCT/US00/04414
; PRIOR FILING DATE: 2000-02-22
; PRIOR APPLICATION NUMBER: US 60/143,048
; PRIOR FILING DATE: 1999-07-07
; PRIOR APPLICATION NUMBER: US 60/145,698
; PRIOR FILING DATE: 1999-07-26
; PRIOR APPLICATION NUMBER: US 60/146,222
; PRIOR FILING DATE: 1999-07-28
; PRIOR APPLICATION NUMBER: PCT/US99/20594
; PRIOR FILING DATE: 1999-09-08
; PRIOR APPLICATION NUMBER: PCT/US99/20944
; PRIOR FILING DATE: 1999-09-13
; PRIOR APPLICATION NUMBER: PCT/US99/21090
; PRIOR FILING DATE: 1999-09-15
; PRIOR APPLICATION NUMBER: PCT/US99/21547

; PRIOR FILING DATE: 1999-09-15
; PRIOR APPLICATION NUMBER: PCT/US99/23089
; PRIOR FILING DATE: 1999-10-05
; PRIOR APPLICATION NUMBER: PCT/US99/28214
; PRIOR FILING DATE: 1999-11-29
; PRIOR APPLICATION NUMBER: PCT/US99/28313
; PRIOR FILING DATE: 1999-11-30
; PRIOR APPLICATION NUMBER: PCT/US99/28564
; PRIOR FILING DATE: 1999-12-02
; PRIOR APPLICATION NUMBER: PCT/US99/28565
; PRIOR FILING DATE: 1999-12-02
; PRIOR APPLICATION NUMBER: PCT/US99/30095
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: PCT/US99/30911
; PRIOR FILING DATE: 1999-12-20
; PRIOR APPLICATION NUMBER: PCT/US99/30999
; PRIOR FILING DATE: 1999-12-20
; PRIOR APPLICATION NUMBER: PCT/US00/00219
; PRIOR FILING DATE: 2000-01-05
; NUMBER OF SEQ ID NOS: 423
; SEQ ID NO 263
; LENGTH: 317
; TYPE: PRT
; ORGANISM: Homo Sapien
US-09-907-794A-263
Query Match 27.9%; Score 544; DB 4; Length 317;
Best Local Similarity 39.3%; Pred. No. 5.4e-42;
Matches 114; Conservative 41; Mismatches 123; Indels 12; Gaps 4;
QY 7 LGPGQLGAVANSYSYLYGLVPSGARGPPYCGRPESARIVGGSNAQPGTWPQVSLHH 66
Db 10 LGGCGCLGTTSLLLASTAILNAARIIPVPACGKPOQLNRVVGEDSTDSEWPIVSIQK 69
QY 67 GGGHICGSLIAPSWVLSAAHCFMTNGTLEPAAEWSVLLGVHSDGPLDGAHTRAVAAIV 126
Db 70 NGTHCAGSLTSSRWITAAHCFKDN--LNKPYLFSVLLGAWQLGNFGRSQKVGVAWVE 127
QY 127 VPANYSQVE--LGADLALIRLASPASLGPAVMPVCLPRASHRFVHGTACWATGWDVQAD 185
Db 128 PHFYVSWKEGACADIALVRLERSIQFSERVLPICLPDASIHLPNTHCWSGSIQGV 187
QY 186 PLPLPWLQVVELRLGATCQCLY---SQGPNLTLQILPGMLCAGYPSGRDTCQGD 242
Db 188 PLPHPQTLQKLKVPIDSEVCSHLYWRGAGQGP-----ITEDMLCAGYLEGERDACLGD 241
QY 243 SGGPLVCEEGRGWFOAGITTSFGCGRNRPGVFTAVATYEAWIREQVMG 292
Db 242 SGGPLMCQVDGAWLLAGLIISWEGCAERNRFGVYISLSAHSRWSVEKIVQG 291

RESULT 9

US-09-905-125A-263
; Sequence 263, Application US/09905125A
; Patent No. 6664376
; GENERAL INFORMATION:
; APPLICANT: Genentech, Inc.
; APPLICANT: Ashkenazi, Avi
; APPLICANT: Botstein, David
; APPLICANT: Desnoyers, Luc
; APPLICANT: Eaton, Dan L.
; APPLICANT: Ferrara, Napoleone
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Fong, Sherman
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerber, Hanspeter
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, A.
; APPLICANT: Godowski, Paul J.
; APPLICANT: Grimaldi, Christopher J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Hillan, Kenneth, J.
; APPLICANT: Kljavin, Ivar J.

APPLICANT: Mather, Jennie P.
APPLICANT: Pan, James
APPLICANT: Paoni, Nicholas F.
APPLICANT: Roy, Margaret Ann
APPLICANT: Stewart, Timothy A.
APPLICANT: Tumas, Daniel
APPLICANT: Williams, P. Mickey
APPLICANT: Wood, William I.
TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
TITLE OF INVENTION: Acids Encoding the Same
FILE REFERENCE: 10466-14
CURRENT APPLICATION NUMBER: US/09/905/125A
CURRENT FILING DATE: 2001-07-12
PRIOR APPLICATION NUMBER: PCT/US00/04414
PRIOR FILING DATE: 2000-02-22
PRIOR APPLICATION NUMBER: US 60/143,048
PRIOR FILING DATE: 1999-07-07
PRIOR APPLICATION NUMBER: US 60/145,698
PRIOR FILING DATE: 1999-07-26
PRIOR APPLICATION NUMBER: US 60/146,222
PRIOR FILING DATE: 1999-07-28
PRIOR APPLICATION NUMBER: PCT/US99/20594
PRIOR FILING DATE: 1999-09-08
PRIOR APPLICATION NUMBER: PCT/US99/20944
PRIOR FILING DATE: 1999-09-13
PRIOR APPLICATION NUMBER: PCT/US99/21090
PRIOR FILING DATE: 1999-09-15
PRIOR APPLICATION NUMBER: PCT/US99/21547
PRIOR FILING DATE: 1999-09-15
PRIOR APPLICATION NUMBER: PCT/US99/23089
PRIOR FILING DATE: 1999-10-05
PRIOR APPLICATION NUMBER: PCT/US99/28214
PRIOR FILING DATE: 1999-11-29
PRIOR APPLICATION NUMBER: PCT/US99/28313
PRIOR FILING DATE: 1999-11-30
PRIOR APPLICATION NUMBER: PCT/US99/28564
PRIOR FILING DATE: 1999-12-02
PRIOR APPLICATION NUMBER: PCT/US99/28565
PRIOR FILING DATE: 1999-12-02
PRIOR APPLICATION NUMBER: PCT/US99/30095
PRIOR FILING DATE: 1999-12-16
PRIOR APPLICATION NUMBER: PCT/US99/30911
PRIOR FILING DATE: 1999-12-20
PRIOR APPLICATION NUMBER: PCT/US99/30999
PRIOR FILING DATE: 1999-12-20

Query Match 27.9%; Score 544; DB 4; Length 317;
Best Local Similarity 39.3%; Pred. No. 5.4e-42;
Matches 114; Conservative 41; Mismatches 123; Indels 12; Gaps 4;
7 LGPGQIGAVANSYSILYGLVSPGARGPPYCGRPESARIIVCGSNAQCTWQVSLHH 66
10 LGGCGLGFTSLLLASTATINAAIRPVPPACGKPOOLNRVGGEDSTDSEWPWISIOK 69
67 GGGHICGSLIAPSWLSAAHCFMTNGTLEPAEWSVLLGVHSDGLDGAHTRAAVIV 126
70 NGTHCAGSLTGRSWVITAACHFKN--LNKPYLFVLLGAWQLGNPGRSQRGVANWE 127
127 VPANYSQVE-IGADLALLRLSPASLGPVAVMPCVLPASHRFVHGFTACWATGNGVQDEAD 185
128 PHPVYSWKEGACADIALVRLERSIQFSERVLPICLPDASIHLPENTHCWISGWGSIQDGV 187
186 PLPLPWVQLVEVRLLEGATCCQCLY----SQGPPNLTQILPGMLCAGYEGGRDTCQD 242
188 PLPHPTQTLQKLVPIIDSEVCSHLYNRGAGQGP-----ITEDMLCAGYLEGERDACLGD 241

QY 243 SGGPLVCEGGRWFQAGITSGFGCCRRNRPVGTAVATVATYEAIRBOVMG 292
Db 242 SGGPLMCQVDGAWLLAGIISWEGCAERNRPGVVVISLSAHSRWVEKIVQG 291
RESULT 10
US-09-902-775A-263
Sequence 263, Application US/09902775A
Patent No. 6686451
GENERAL INFORMATION:
APPLICANT: Genentech, Inc.
APPLICANT: Ashkenazi, Avi
APPLICANT: Botstein, David
APPLICANT: Desnoyers, Luc
APPLICANT: Eaton, Dan L.
APPLICANT: Ferrara, Napoleone
APPLICANT: Filvaroff, Ellen
APPLICANT: Fong, Sherman
APPLICANT: Gao, Wei-Qiang
APPLICANT: Gerber, Hanspeter
APPLICANT: Geritsen, Mary E.
APPLICANT: Goddard, A.
APPLICANT: Godowski, Paul J.
APPLICANT: Grimaldi, Christopher J.
APPLICANT: Gurney, Austin L.
APPLICANT: Hillan, Kenneth, J.
APPLICANT: Kljavin, Ivar J.
APPLICANT: Mather, Jennie P.
APPLICANT: Pan, James
APPLICANT: Paoni, Nicholas F.
APPLICANT: Roy, Margaret Ann
APPLICANT: Stewart, Timothy A.
APPLICANT: Tumas, Daniel
APPLICANT: Williams, P. Mickey
APPLICANT: Wood, William I.
TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
TITLE OF INVENTION: Acids Encoding the Same
FILE REFERENCE: 10466-14
CURRENT APPLICATION NUMBER: US/09/902,775A
CURRENT FILING DATE: 2001-07-10
PRIOR APPLICATION NUMBER: PCT/US00/04414
PRIOR FILING DATE: 2000-02-22
PRIOR APPLICATION NUMBER: US 60/143,048
PRIOR FILING DATE: 1999-07-07
PRIOR APPLICATION NUMBER: US 60/145,698
PRIOR FILING DATE: 1999-07-26
PRIOR APPLICATION NUMBER: US 60/146,222
PRIOR FILING DATE: 1999-07-28
PRIOR APPLICATION NUMBER: PCT/US99/20594
PRIOR FILING DATE: 1999-09-08
PRIOR APPLICATION NUMBER: PCT/US99/20944
PRIOR FILING DATE: 1999-09-13
PRIOR APPLICATION NUMBER: PCT/US99/21090
PRIOR FILING DATE: 1999-09-15
PRIOR APPLICATION NUMBER: PCT/US99/21547
PRIOR FILING DATE: 1999-09-15
PRIOR APPLICATION NUMBER: PCT/US99/23089
PRIOR FILING DATE: 1999-10-05
PRIOR APPLICATION NUMBER: PCT/US99/28214
PRIOR FILING DATE: 1999-11-29
PRIOR APPLICATION NUMBER: PCT/US99/28313
PRIOR FILING DATE: 1999-11-30
PRIOR APPLICATION NUMBER: PCT/US99/28564
PRIOR FILING DATE: 1999-12-02
PRIOR APPLICATION NUMBER: PCT/US99/28565
PRIOR FILING DATE: 1999-12-02
PRIOR APPLICATION NUMBER: PCT/US99/30095
PRIOR FILING DATE: 1999-12-16
PRIOR APPLICATION NUMBER: PCT/US99/30911
PRIOR FILING DATE: 1999-12-20
PRIOR APPLICATION NUMBER: PCT/US99/30999
PRIOR FILING DATE: 1999-12-20


```
; PRIOR FILING DATE: 1999-10-20
;   1998-02-20
; NUMBER OF SEQ ID NOS: 98
; SEQ ID NO 2
; LENGTH: 855
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: TADG-15
US-09-654-600A-2

Query Match          27.6%; Score 540; DB 4; Length 855;
Best Local Similarity 40.2%; Pred. No. 4.8e-41;
Matches 113; Conservative 39; Mismatches 95; Indels 34; Gaps 7;

QY      32 ARGPPVCGRREP-----SARIVGSNAQPCTPWQVSLHH-GGGHI 71
       ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db      581 SKNPECDKEDCSGDSDXDCDGLSFTRQARVVGTDADEGWPQVSLHALGGCHI 640

QY      72 CGSGLIAPSWLGAHCFMNG--TLSPAABWSVLGLVHSQ-DPLDGATRAVAALVVP 128
       |||||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db      641 CGASLIAPNLWGAACIIIDRRFRYSDDPTQTWTAFGLHDQSRSAPGVERRLKRIISH 700

QY      129 ANYSQVELGADLALLRLASLASLPAPVCLPEASHRFVHGTCWATGMGDVQEADPLP 188
       ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db      701 PFNDFTFDYDIALLLEKEAAYSSMRPICLPDASHVFPAGKAIAIWVGHTQYGGTGA 760

QY      189 LPWLVQVEILRLIGEATCCQLYSQPGPFNLTLQLPGMLCAGYPEGRDRDTCCGDSGGPL- 247
       ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db      761 L-L-LLQGEIRVINQTTENLLPQ-----QITPRMCVGFLLSGGVDSQQGDSGGPLS 810

QY      248 VCBEGGWFQAGITSFGGCGRRNRPGVFTAVATYEAWIRE 288
       ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db      811 SVRADGRIFQAGVSVWGDCGAORKNPGYTRLPLFRDWIKE 851
```

RESULT 14

US-09-008-271A-3

Sequence 3, Application US/09008271A
Patent No. 6203979
GENERAL INFORMATION:
APPLICANT: Hillman, Jennifer L.
Yue, Henry
Guegler, Karl J.
Corley, Neil C.
Tang, Tom Y.
Shah, Purvi
TITLE OF INVENTION: HUMAN PROTEASE MOLECULES
NUMBER OF SEQUENCES: 24
CORRESPONDENCE ADDRESS:
ADDRESSEE: Incyte Pharmaceuticals, Inc.
STREET: 3174 Porter Dr.
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94304

```

;
; TELEPHONE: 650-855-0555
; TELEFAX: 650-845-4166
; INFORMATION FOR SEQ ID: NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 314 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: PROSTUT03
; CLONE: 789927
; SEQUENCE DESCRIPTION: SEQ ID NO: 3 :
US-09-008-271A-3

Query Match 27.4%; Score 534.5; DB 3; Length 314;
Best Local Similarity 39.0%; Pred. No. 4e-41;
Matches 110; Conservative 53; Mismatches 100; Indels 19; Gaps 8;

QY 29 SGPARGPYCGRPEPSARIVGGSNAOPGWQVSLHHGGGHHCGSLIAPSWVLSAHC 88
Db 26 AAPLSGP--CGRRVITSRIVGGEDAEIGRWQOGLRSLNDSHVCGVSLLSHRWALTAHC 83
QY 89 FMTNGTLEPAAEWSVLG-VHSQDG--ELDGAHTR-AVAAIIVPANKYSQVELG---ADLA 141
Db 84 FETYSDLSPSGRWVQFGLTSPFSFWSLQAYTRYFVSNILSPRY---LGNSPYDIA 139
QY 142 LRLASPASLGPAWVPVCLIPRASHRFVHGHTACWATGMDGVQOEADPLPLPWVLQVEVRL 201
Db 140 LVKLSPAVITYTHIQICLQASTFEFENRTDCWVTGWTGXYIKEDALESPHTLQEVQVAIL 199
QY 202 GEATCCCLQSQPGPNLTLLQILPGLCAGYPEGRDRDTCQDGSQGLPVCEEGGRWFQAGIT 261
Db 200 NNSMCHLFLK---YFPRKDI FGMVVCAGNAQGGKDACFGDSGGLACNKNGLWYQIGV 256
QY 262 SFQGGCRRNRPGVTTAVATYFAWIRE---QVMGSEFGPAPP 300
Db 257 SWGVGCGRRNRPGVYTNISHHFEWIIQKLMAQSGMSQDPDPSWP 298

```

RESULT 15

US-09-907-794A-257
; Sequence 257, Application US/09907794A
; Patent No. 6635468
; GENERAL INFORMATION:
; APPLICANT: Genentech, Inc.
; APPLICANT: Ashkenazi, Avi
; APPLICANT: Botstein, David
; APPLICANT: Desnoyers, Luc
; APPLICANT: Eaton, Dan L.
; APPLICANT: Ferrara, Napoleone
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Fong, Sherman
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerber, Hanspeter
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, A.
; APPLICANT: Godowski, Paul J.
; APPLICANT: Grimaldi, Christopher J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Hillan, Kenneth, J.
; APPLICANT: Kijavín, Ivar J.
; APPLICANT: Mather, Jennie P.
; APPLICANT: Pan, James
; APPLICANT: Paoni, Nicholas F.
; APPLICANT: Roy, Margaret Ann
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Williams, P. Mickey
; APPLICANT: Wood, William, I.
; TITLE OF INVENTION: Secreted and Trans-
; TITLE OF INVENTION: Acids Encoding t
; FILE REFERENCE: 10466-14
; CURRENT APPLICATION NUMBER: US/09/907

```

; CURRENT FILING DATE: 2001-07-17
; PRIOR APPLICATION NUMBER: PCT/US00/04414
; FILING DATE: 2000-02-22
; PRIOR APPLICATION NUMBER: US 60/143,048
; FILING DATE: 1999-07-07
; PRIOR APPLICATION NUMBER: US 60/145,698
; FILING DATE: 1999-07-26
; PRIOR APPLICATION NUMBER: US 60/146,222
; FILING DATE: 1999-07-28
; PRIOR APPLICATION NUMBER: PCT/US99/20594
; FILING DATE: 1999-09-08
; PRIOR APPLICATION NUMBER: PCT/US99/20944
; FILING DATE: 1999-09-13
; PRIOR APPLICATION NUMBER: PCT/US99/21090
; FILING DATE: 1999-09-15
; PRIOR APPLICATION NUMBER: PCT/US99/21547
; FILING DATE: 1999-09-15
; PRIOR APPLICATION NUMBER: PCT/US99/23089
; FILING DATE: 1999-10-05
; PRIOR APPLICATION NUMBER: PCT/US99/28214
; FILING DATE: 1999-11-29
; PRIOR APPLICATION NUMBER: PCT/US99/28313
; FILING DATE: 1999-11-30
; PRIOR APPLICATION NUMBER: PCT/US99/28564
; FILING DATE: 1999-12-02
; PRIOR APPLICATION NUMBER: PCT/US99/28565
; FILING DATE: 1999-12-02
; PRIOR APPLICATION NUMBER: PCT/US99/30095
; FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: PCT/US99/30911
; FILING DATE: 1999-12-20
; PRIOR APPLICATION NUMBER: PCT/US99/30999
; FILING DATE: 1999-12-20
; PRIOR APPLICATION NUMBER: PCT/US00/00219
; FILING DATE: 2000-01-05
; NUMBER OF SEQ ID NOS: 423
; SEQ ID NO 257
; LENGTH: 314
; TYPE: PRT
; ORGANISM: Homo Sapien
US-09-907-794A-257

Query Match      27.4%; Score 534.5; DB 4; Length 314;
Best Local Similarity 39.0%; Pred. No. 4e-41;
Matches 110; Conservative 53; Mismatches 100; Indels 19; Gaps 8;

QY 29 SGARGPPYCGREPSARIYGGNAQGTWPQVSLHHGGHICGGSLIAPSWLSAAHC 88
DB 26 AAPLSGP--CGRRVITSRIVGGDAELGRWPQCSLLWDSHVCGVSLLSHRWALTAHC 83
QY 89 FMTNGTLEPAAEWSVLIG--VHSQDG--PLDGAHTR-AVAAIIVVPANYSQVELG---ADLA 141
DB 84 FETYSLSDFSGWMVQFGLTSMPSFWSLQAYTRYFVSNLYLSPRY---LGNSPYDIA 139
QY 142 LLRLASPASLGPAPVPCVLPASHRFVHGTHACWATGWDVQVADPLPWPVLQVELRL 201
DB 140 LVKLSAPVTYTKHIQICLOASTFEFFENRTDCWITGNGYIKEDFALPSPHTLQEVQVAIL 199
QY 202 GEATCCQLYSGPGFNLTLQILPGMLCAGYEGRRDTCQDSDGGPLVCEEGGRWFQAGIT 261
DB 200 NNSMCHLFLK--YSFRKDIFGDMVAGNAQGGKDACFGDSGGPLACNKNGLWYQLGVV 256
QY 262 SFEGCGRRNRPGVFTAVATYEAWIRE---QVMGSEPGAPF 300
DB 257 SWGVGCGRPNEGVTYINSHHFEMIQLMAQSGMSQDPDSWP 298

```

Search completed: February 25, 2004, 15:51:28
Job time : 46 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 25, 2004, 15:49:38 ; Search time 641 Seconds
(without alignments)
117.600 Million cell updates/sec

Title: US-10-037-417-46

Perfect score: 1953

Sequence: 1 MAQKGVLGQGLGAVANS...TKSLVLPWLSPHSLGLGWGF 357

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 809742 seqs, 211153259 residues

Total number of hits satisfying chosen parameters: 809742

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- Published Applications AA: *
- 1: /cgn2_6/prodata/1/pubpaa/US07_PUBCOMB.pep.*
 - 2: /cgn2_6/prodata/1/pubpaa/PCT_NEW_PUB.pep.*
 - 3: /cgn2_6/prodata/1/pubpaa/US05_NEW_PUB.pep.*
 - 4: /cgn2_6/prodata/1/pubpaa/US06_PUBCOMB.pep.*
 - 5: /cgn2_6/prodata/1/pubpaa/US07_NEW_PUB.pep.*
 - 6: /cgn2_6/prodata/1/pubpaa/PCTUS_PUBCOMB.pep.*
 - 7: /cgn2_6/prodata/1/pubpaa/US08_NEW_PUB.pep.*
 - 8: /cgn2_6/prodata/1/pubpaa/US08_PUBCOMB.pep.*
 - 9: /cgn2_6/prodata/1/pubpaa/US09A_PUBCOMB.pep.*
 - 10: /cgn2_6/prodata/1/pubpaa/US09B_PUBCOMB.pep.*
 - 11: /cgn2_6/prodata/1/pubpaa/US09C_PUBCOMB.pep.*
 - 12: /cgn2_6/prodata/1/pubpaa/US09_NEW_PUB.pep.*
 - 13: /cgn2_6/prodata/1/pubpaa/US10A_PUBCOMB.pep.*
 - 14: /cgn2_6/prodata/1/pubpaa/US10B_PUBCOMB.pep.*
 - 15: /cgn2_6/prodata/1/pubpaa/US10C_PUBCOMB.pep.*
 - 16: /cgn2_6/prodata/1/pubpaa/US10_NEW_PUB.pep.*
 - 17: /cgn2_6/prodata/1/pubpaa/US60_NEW_PUB.pep.*
 - 18: /cgn2_6/prodata/1/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1505	77.1	818	9	US-09-888-615-111
2	690	35.3	343	9	Sequence 111, App
3	690	35.3	343	14	Sequence 2, Appli
4	690	35.3	343	14	Sequence 86, Appl
5	690	35.3	343	15	Sequence 262, App
6	627.5	32.1	339	14	Sequence 125, App
7	621	31.8	342	15	Sequence 2, Appli
8	620	31.7	342	15	Sequence 86, Appl
9	590	30.2	280	15	Sequence 87, Appl
10	590	30.2	284	13	Sequence 26, Appl
11	590	30.2	284	13	Sequence 7, Appli
12	590	30.2	284	13	Sequence 7, Appli
13	583.5	29.9	285	15	Sequence 89, Appl
14	582.5	29.8	346	16	Sequence 8, Appli
15	575.5	29.5	255	14	Sequence 36, Appl

16	567	29.0	389	15	US-10-074-978A-219	Sequence 219, App
17	563	28.8	290	11	US-09-833-245-1294	Sequence 1294, Ap
18	563	28.8	290	13	US-10-040-655-7	Sequence 7, Appli
19	563	28.8	290	13	US-10-041-006A-7	Sequence 7, Appli
20	563	28.8	290	14	US-10-028-072-222	Sequence 222, App
21	563	28.8	290	14	US-10-121-049-222	Sequence 222, App
22	563	28.8	290	14	US-10-123-904-222	Sequence 222, App
23	563	28.8	290	14	US-10-140-470-222	Sequence 222, App
24	563	28.8	290	14	US-10-175-746-222	Sequence 222, App
25	563	28.8	290	14	US-10-176-918-222	Sequence 222, App
26	563	28.8	290	14	US-10-176-921-222	Sequence 222, App
27	563	28.8	290	14	US-10-137-865-222	Sequence 222, App
28	563	28.8	290	14	US-10-140-474-222	Sequence 222, App
29	563	28.8	290	14	US-10-142-431-222	Sequence 222, App
30	563	28.8	290	14	US-10-143-114-222	Sequence 222, App
31	563	28.8	290	14	US-10-140-002-222	Sequence 222, App
32	563	28.8	290	14	US-10-142-419-222	Sequence 222, App
33	563	28.8	290	14	US-10-123-262-222	Sequence 222, App
34	563	28.8	290	14	US-10-142-423-222	Sequence 222, App
35	563	28.8	290	14	US-10-121-050-222	Sequence 222, App
36	563	28.8	290	14	US-10-141-755-222	Sequence 222, App
37	563	28.8	290	14	US-10-143-032-222	Sequence 222, App
38	563	28.8	290	14	US-10-123-108-222	Sequence 222, App
39	563	28.8	290	14	US-10-123-236-222	Sequence 222, App
40	563	28.8	290	14	US-10-123-261-222	Sequence 222, App
41	563	28.8	290	14	US-10-140-921-222	Sequence 222, App
42	563	28.8	290	14	US-10-140-928-222	Sequence 222, App
43	563	28.8	290	14	US-10-121-045-222	Sequence 222, App
44	563	28.8	290	14	US-10-123-292-222	Sequence 222, App
45	563	28.8	290	14	US-10-123-903-222	Sequence 222, App

ALIGNMENTS

RESULT 1

US-09-888-615-111
; Sequence 111, Application US/09888615
; Patent No. US20020064856A1
; GENERAL INFORMATION:
; APPLICANT: PLOWMAN, GREGORY
; APPLICANT: WHYTE, DAVID
; APPLICANT: CAENEPEEL, SEAN
; APPLICANT: CHARYDCZAK, GLEN
; APPLICANT: MANNING, GERARD
; APPLICANT: SUDARSANAM, SUCHA
; TITLE OF INVENTION: NOVEL PROTEASES
; FILE REFERENCE: 038602/1214
; CURRENT APPLICATION NUMBER: US/09/888,615
; PRIOR FILING DATE: 2001-06-26
; PRIOR APPLICATION NUMBER: 60/214,047
; PRIOR FILING DATE: 2000-06-26
; NUMBER OF SEQ ID NOS: 150
; SOFTWARE: Patent in Ver. 2.1
; SEQ ID NO 111
; LENGTH: 818
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-888-615-111

Query Match 77.1%; Score 1505; DB 9; Length 818;
Best Local Similarity 100.0%; Pred. No. 6.1e-122;
Matches 272; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy	38	CGPEPSARIVGSSNAQPGTWQVSLHHGGGHHCGSLIAPSWVLSAAHCFMTNGTLEP	97
Db	38	CGPEPSARIVGSSNAQPGTWQVSLHHGGGHHCGSLIAPSWVLSAAHCFMTNGTLEP	97
Qy	98	AAEWSVLLGVHSDGFLDGAHTRAVAAIVVPANYSOVELGADIALRLASPSLGFAPVWP	157
Db	98	AAEWSVLLGVHSDGFLDGAHTRAVAAIVVPANYSOVELGADIALRLASPSLGFAPVWP	157
Qy	158	VCLPRASHRFVHGTAACWATGWGDVQEADELPLPWLQVEVLRLLGEATCQCLYQSGPFN	217

```

; NUMBER OF SEQ ID NOS: 112
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO: 86
; LENGTH: 343
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-176-847-86

Query Match      35.3%; Score 690; DB 14; Length 343;
Best Local Similarity 43.6%; Pred. No. 1.1e-51;
Matches 156; Conservative 49; Mismatches 129; Indels 24; Gaps 10;

1 MAQKVLGPGQLGAVANSYSY-GLVPSGP-----ARGPPYCGRPEPSARIVGGSSNAQP 55
1 MAQKVLGPGQLGAVA-----ILYLGLLRSGTGAEGAEAP--CG-VAPQARITGSSAVA 53
56 GTWPMOVSLLHHGGHICGSLIAPSWLSAAHCFMTNGTLEPAAEWSVLLGVHSQDGLD 115
54 GQWPQVSIITVEGVHVCGGLVSEQWVLSAACHFPSEHHKE---AYEVKLGAHQDLSYSE 110
116 GAHTRAVAAIVVPANYSQVELGADLALLRLASPASLGPVWPVCLPRASHRFVHGTACWA 175
111 DAKVSTLKDIIIPHPSYLQEGSQGDIALQLSRPITFSYIRPICLPAANASFPNGLHCTV 170
176 TGMGDVQADPLPLPWLQVLELLEGEATCCQLYSQPGPFNLTLQLPGMLCAGYPEGR 235
171 TGMGHVAPSVSLTPKPLQQLQLEVLISRETNCNLYNIDAKPEEPHFVQEDMVCAGYVEGG 230
236 RDTCCGDSGGPLVCEEGGRWFQAGITSGFCGGRNRPGVFTAVATYEAWIREQVMGSEP 295
231 KDAQCGDSGGPLSCPVEGLWYLTGIVSWGDCAGARNRPGVYTLASSYASWISQKV--TEL 288
296 GPAPFTQPKTQSD---CLHQTAFIDS-ARILLRPLSHISVGVSTGTSKSLVLPWLSPH 349
289 QPRVVPQTQESQPSDNLGSHLAFSSAPAQGLLRPIILFLPLGLALG---LLSPWLSEH 343

RESULT 4
US-10-097-340-262
Sequence 262, Application US/10097340
Publication No. US20030087250A1
GENERAL INFORMATION:
APPLICANT: John MONAHAN
APPLICANT: Manjula GANNAVAPAPU
APPLICANT: Sebastian HOERSCH
APPLICANT: Shubhangi KAMATKAR
APPLICANT: Steve G. KOVATS
APPLICANT: Rachel B. MEYERS
APPLICANT: Michael MORRISSEY
APPLICANT: Peter OLANDT
APPLICANT: Ami SEN
APPLICANT: Gordon B. MILLS
APPLICANT: Robert C. BAST, Jr.
APPLICANT: Karen LU
APPLICANT: Rosemarie SCHMANDT
APPLICANT: Xumel ZHAO
APPLICANT: Karen GLATT
TITLE OF INVENTION: Nucleic Acid Molecules and Proteins For The Identification,
TITLE OF INVENTION: Assessment, Prevention, and Therapy Of Ovarian Cancer
FILE REFERENCE: MRI-030
CURRENT APPLICATION NUMBER: US/10/097,340
CURRENT FILING DATE: 2002-03-14
PRIOR APPLICATION NUMBER: 60/276,025
PRIOR FILING DATE: 2001-03-14
PRIOR APPLICATION NUMBER: 60/325,149
PRIOR FILING DATE: 2001-09-26
PRIOR APPLICATION NUMBER: 60/276,026
PRIOR FILING DATE: 2001-03-14
PRIOR APPLICATION NUMBER: 60/324,967
PRIOR FILING DATE: 2001/09/26
PRIOR APPLICATION NUMBER: 60/311,732
PRIOR FILING DATE: 2001-08-10

```

```

; NUMBER OF SEQ ID NOS: 112
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO: 86
; LENGTH: 343
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-176-847-86

Query Match      35.3%; Score 690; DB 14; Length 343;
Best Local Similarity 43.6%; Pred. No. 1.1e-51;
Matches 156; Conservative 49; Mismatches 129; Indels 24; Gaps 10;

1 MAQKVLGPGQLGAVANSYSY-GLVPSGP-----ARGPPYCGRPEPSARIVGGSSNAQP 55
1 MAQKVLGPGQLGAVA-----ILYLGLLRSGTGAEGAEAP--CG-VAPQARITGSSAVA 53
56 GTWPMOVSLLHHGGHICGSLIAPSWLSAAHCFMTNGTLEPAAEWSVLLGVHSQDGLD 115
54 GQWPQVSIITVEGVHVCGGLVSEQWVLSAACHFPSEHHKE---AYEVKLGAHQDLSYSE 110
116 GAHTRAVAAIVVPANYSQVELGADLALLRLASPASLGPVWPVCLPRASHRFVHGTACWA 175
111 DAKVSTLKDIIIPHPSYLQEGSQGDIALQLSRPITFSYIRPICLPAANASFPNGLHCTV 170
176 TGMGDVQADPLPLPWLQVLELLEGEATCCQLYSQPGPFNLTLQLPGMLCAGYPEGR 235
171 TGMGHVAPSVSLTPKPLQQLQLEVLISRETNCNLYNIDAKPEEPHFVQEDMVCAGYVEGG 230
236 RDTCCGDSGGPLVCEEGGRWFQAGITSGFCGGRNRPGVFTAVATYEAWIREQVMGSEP 295
231 KDAQCGDSGGPLSCPVEGLWYLTGIVSWGDCAGARNRPGVYTLASSYASWISQKV--TEL 288
296 GPAPFTQPKTQSD---CLHQTAFIDS-ARILLRPLSHISVGVSTGTSKSLVLPWLSPH 349
289 QPRVVPQTQESQPSDNLGSHLAFSSAPAQGLLRPIILFLPLGLALG---LLSPWLSEH 343

RESULT 3
US-10-176-847-86
Sequence 86, Application US/10176847
Publication No. US20030068636A1
GENERAL INFORMATION:
APPLICANT: Veiby, Petter Ole
TITLE OF INVENTION: COMPOSITIONS, KITS, AND METHODS FOR
TITLE OF INVENTION: IDENTIFICATION, ASSESSMENT, PREVENTION, AND THERAPY OF BREAST
TITLE OF INVENTION: IDENTIFICATION, ASSESSMENT, PREVENTION, AND THERAPY OF BREAST
FILE REFERENCE: MRI-039
CURRENT APPLICATION NUMBER: US/10/176,847
CURRENT FILING DATE: 2002-06-21

```

; PRIOR APPLICATION NUMBER: 60/325,102
 ; PRIOR FILING DATE: 2001-09-26
 ; PRIOR APPLICATION NUMBER: 60/323,580
 ; PRIOR FILING DATE: 2001-09-19
 ; NUMBER OF SEQ ID NOS: 363
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 262
 ; LENGTH: 343
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 US-10-097-340-262

Query Match 35.3%; Score 690; DB 14; Length 343;
 Best Local Similarity 43.6%; Pred. No. 1.1e-51;
 Matches 156; Conservative 49; Mismatches 129; Indels 24; Gaps 10;
 QY 1 MAQKGVLPQGLGAVANSYSYLY-GLVPSGP-ARGPPYCGRPESARIVGSGNAQP 55
 DB 1 MAQKGVLPQGLGAVA-ILLYLGLLSRGTGAEGAEAP--CG-VAPOARITGSSAVA 53
 QY 56 GTPWQVSLHGGGHI CGSLIAPSWLSAAHCFWNTGLTLEPAAEWSVLLGVHSDGPLD 115
 DB 54 GOWPQVSIITTEGVHVCGLSVSEQWLSAAHCFPSEHKE---AYEVKLGHAHQDLSYSE 110
 QY 116 GAHTRAVAAIVVPANYSQVELGADLALLRLASPSLGPVAVMPVCLPRASHRFVHGTA 175
 DB 111 DAKVSTLKDIIIPHSYVQLQSGQDIALQLSRITFSYRIRPICLPAAANASFPNGLHCTV 170
 QY 176 TQMGVDQVQADPLPLPWLQVELELRLGEATCCQLYSQPGFNLTLQILPGMLCAGYPEG 235
 DB 171 TQMGHVAPSVLLTPKPLQQLVPLISRETCLNLYNIDAKPEEPHFVQEDMVCAGYVEGG 230
 QY 236 RDTQCGDSGGLPVCBEGGRWFQAGITSFGCGRRNRPVFTAVATYEAWAREQVWGSEP 295
 DB 231 KDACQDSGGLPVCBEGGLWYLTGIVSWGDCAGARNRPGVYTLASSYASWIOSKV--TEL 288
 QY 296 GRAFPTQPKTQSD---CLHQTAFLDS-ARILLRPLSHISVGVSTGKSLVPLWLSPH 349
 DB 289 QPRVVPQTQESQPDNLGSHLAFSSAPAGQLLRPLFLPLGLAUG---LLSPWLSEH 343

RESULT 5
 US-10-074-566-125
 ; Sequence 125, Application US/10074566
 ; Publication No. US20030207348A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Shimkets, Richard A.
 ; APPLICANT: Fernandes, Elma R.
 ; APPLICANT: Li, Li
 ; APPLICANT: Gorman, Linda
 ; APPLICANT: Gusev, Vladimir Y.
 ; APPLICANT: Padigaru, Muralidhara
 ; APPLICANT: Patturajan, Meera
 ; APPLICANT: Shenoy, Suresh G.
 ; APPLICANT: Spytek, Kimberly A.
 ; TITLE OF INVENTION: Polypeptides and Polynucleotides Encoding Same
 ; FILE REFERENCE: 15966-556 CIP1
 ; CURRENT APPLICATION NUMBER: US/10/074,566
 ; CURRENT FILING DATE: 2002-02-13
 ; PRIOR APPLICATION NUMBER: 09/619,252
 ; PRIOR FILING DATE: 2000-07-19
 ; PRIOR APPLICATION NUMBER: 60/144,722
 ; PRIOR FILING DATE: 1999-07-20
 ; PRIOR APPLICATION NUMBER: 60/167,785
 ; PRIOR FILING DATE: 1999-11-29
 ; PRIOR APPLICATION NUMBER: 60/276,994
 ; PRIOR FILING DATE: 2001-03-19
 ; PRIOR APPLICATION NUMBER: 60/280,898
 ; PRIOR FILING DATE: 2001-04-02
 ; PRIOR APPLICATION NUMBER: 60/332,241
 ; PRIOR FILING DATE: 2001-11-14
 ; PRIOR APPLICATION NUMBER: 60/288,062
 ; PRIOR FILING DATE: 2001-05-02

; PRIOR APPLICATION NUMBER: 60/291,766
 ; PRIOR FILING DATE: 2001-05-17
 ; PRIOR APPLICATION NUMBER: 60/314,007
 ; PRIOR FILING DATE: 2001-08-21
 ; NUMBER OF SEQ ID NOS: 132
 ; SOFTWARE: PatentIn Ver. 2.1
 ; SEQ ID NO 125
 ; LENGTH: 343
 ; TYPE: PRT
 ; ORGANISM: human
 US-10-074-566-125

Query Match 35.3%; Score 690; DB 15; Length 343;
 Best Local Similarity 43.6%; Pred. No. 1.1e-51;
 Matches 156; Conservative 49; Mismatches 129; Indels 24; Gaps 10;
 QY 1 MAQKGVLPQGLGAVANSYSYLY-GLVPSGP-ARGPPYCGRPESARIVGSGNAQP 55
 DB 1 MAQKGVLPQGLGAVA-ILLYLGLLSRGTGAEGAEAP--CG-VAPOARITGSSAVA 53
 QY 56 GTPWQVSLHGGGHI CGSLIAPSWLSAAHCFWNTGLTLEPAAEWSVLLGVHSDGPLD 115
 DB 54 GOWPQVSIITTEGVHVCGLSVSEQWLSAAHCFPSEHKE---AYEVKLGHAHQDLSYSE 110
 QY 116 GAHTRAVAAIVVPANYSQVELGADLALLRLASPSLGPVAVMPVCLPRASHRFVHGTA 175
 DB 111 DAKVSTLKDIIIPHSYVQLQSGQDIALQLSRITFSYRIRPICLPAAANASFPNGLHCTV 170
 QY 176 TQMGVDQVQADPLPLPWLQVELELRLGEATCCQLYSQPGFNLTLQILPGMLCAGYPEG 235
 DB 171 TQMGHVAPSVLLTPKPLQQLVPLISRETCLNLYNIDAKPEEPHFVQEDMVCAGYVEGG 230
 QY 236 RDTQCGDSGGLPVCBEGGRWFQAGITSFGCGRRNRPVFTAVATYEAWAREQVWGSEP 295
 DB 231 KDACQDSGGLPVCBEGGLWYLTGIVSWGDCAGARNRPGVYTLASSYASWIOSKV--TEL 288
 QY 296 GRAFPTQPKTQSD---CLHQTAFLDS-ARILLRPLSHISVGVSTGKSLVPLWLSPH 349
 DB 289 QPRVVPQTQESQPDNLGSHLAFSSAPAGQLLRPLFLPLGLAUG---LLSPWLSEH 343

RESULT 6
 US-10-109-616-2
 ; Sequence 2, Application US/10109616
 ; Publication No. US2003016748A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Allen, Keith D.
 ; TITLE OF INVENTION: TRANSGENIC MICE CONTAINING CHANNEL
 ; TITLE OF INVENTION: ACTIVATING PROTEASE 1 (CAPI) GENE DISRUPTIONS
 ; FILE REFERENCE: R-490
 ; CURRENT APPLICATION NUMBER: US/10/109,616
 ; CURRENT FILING DATE: 2002-03-28
 ; PRIOR APPLICATION NUMBER: US 60/280,509
 ; PRIOR FILING DATE: 2001-03-29
 ; PRIOR APPLICATION NUMBER: US 60/311,055
 ; PRIOR FILING DATE: 2001-08-08
 ; NUMBER OF SEQ ID NOS: 4
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 2
 ; LENGTH: 339
 ; TYPE: PRT
 ; ORGANISM: Mus musculus
 US-10-109-616-2

Query Match 32.1%; Score 627.5; DB 14; Length 339;
 Best Local Similarity 41.4%; Pred. No. 2.9e-46;
 Matches 144; Conservative 48; Mismatches 139; Indels 17; Gaps 9;
 QY 1 MAQKGVLPQGLGAVANSYSYLYGLVPSG-PARG-PPYCGRP-EPASARIVGSGNAQP 57
 DB 1 MAPRVGLGLGLEAVT---ILLGLLQSGIRADGTETASCGAVTQP--RITGGGSAKPG 55
 QY 58 WPWQVSLHGGGHI CGSLIAPSWLSAAHCFWNTGLTLEPAAEWSVLLGVHSDGPLDGA 117

Db 56 WPMQVSTYDGNHVGCGSLVSNKMWVSAHCFPREHRE---AYEVKLGHQHDSYNDT 112
Qy 118 HTRAVAAIVVPANYSQVELGADLALLRLASPAISGPAVWPVCLPRASHRFVHGTCWATG 177
Db 113 VVHTVAQIITHSSYREEGSQDIAFIRLSBPVTSRVRIRICLPAANASFPNGLHCTVTG 172
Qy 178 WGDVQEAADPLPLPWVLQVEVRLILGEATCOCLYSQPGFNLTLQILPGMLCAGYPEGRRD 237
Db 173 WGHVAPSVSLQTPRPLQOQLEVPILSRFCSCLYNINAVPEEPHTIQDMLCAGYVKGKD 232
Qy 238 TCQDGGGPLYCEEGGRWFQAGITSFGCGRRNRPGVFTAVATYEAWIREQVMSBPFP 297
Db 233 ACQDGGGPLYCEEGGRWFQAGITSFGCGRRNRPGVFTAVATYEAWIREQVMSBPFP 297
Qy 298 AFPTQPKTQSD---CLHQTAF-LDSARILLRPLSHISVGVSTGKSL 341
Db 291 RVVPQTQESQPDGHLGNHHPVFSSAAAPKLLRPVLFPLGLTGLLSL 338

RESULT 7
US-10-051-874-86
; Sequence 86, Application US/10051874
; Publication No. US20040005557A1
; GENERAL INFORMATION:
; APPLICANT: Padigaru, Muralidhara
; APPLICANT: Alsobrook II, John P
; APPLICANT: Colman, Steven D
; APPLICANT: Spytek, Kimberly A
; APPLICANT: Boldog, Ferenc
; APPLICANT: Vernet, Corine AM
; APPLICANT: Li, Li
; APPLICANT: Shency, Suresh G
; APPLICANT: Casman, Stacie J
; APPLICANT: Guo, Xiaojia Sasha
; APPLICANT: Edinger, Shlomit R
; APPLICANT: MacDougall, John R
; APPLICANT: Malyankar, Uriel M
; APPLICANT: Patturajan, Meera
; APPLICANT: Shimkets, Richard A
; APPLICANT: Pena, Carol EA
; APPLICANT: Baumgartner, Jason C
; APPLICANT: Herrman, John L
; APPLICANT: Peyman, John A
; APPLICANT: Gorman, Linda
; APPLICANT: Mezes, Peter D
; APPLICANT: Kekuda, Ramesh
; APPLICANT: Taupier Jr, Raymond J
; APPLICANT: Gerlach, Valerie
; APPLICANT: Grosse, William M
; APPLICANT: Liu, Xiaohong
; APPLICANT: Ellerman, Karen
; APPLICANT: Rothenberg, Mark
; APPLICANT: Stone, David J
; APPLICANT: Burgess, Catherine E
; APPLICANT: Burgess, Catherine E
; TITLE OF INVENTION: PROTEINS, POLYNUCLEOTIDES ENCODING THEM AND METHODS OF
; TITLE OF INVENTION: USING THE SAME
; FILE REFERENCE: 21402-245
; CURRENT APPLICATION NUMBER: US/10/051,874
; CURRENT FILING DATE: 2002-09-25
; PRIOR APPLICATION NUMBER: 60/268,595
; PRIOR FILING DATE: 2001-02-14
; PRIOR APPLICATION NUMBER: 60/325,306
; PRIOR FILING DATE: 2001-09-27
; PRIOR APPLICATION NUMBER: 60/262,587
; PRIOR FILING DATE: 2001-01-18
; PRIOR APPLICATION NUMBER: 60/272,409

Query Match 31.8%; Score 621; DB 15; Length 342;
Best Local Similarity 40.8%; Pred. No. 1.1e-45;
Matches 144; Conservative 50; Mismatches 139; Indels 20; Gaps 9;
Qy 1 MAQGVILGPGQLGAVANSYSYGLVPS--GPARGPPYCGRP--EPSARIYVGGSSNAQPGT 57
Db 1 MALRVGLGLGQLEALF---VLLIGLQSRIGADGTREASCGAVIQP--RITGGGSAKPGQ 55
Qy 58 WPMQVSLHHGGHICGGSLIAPSWVLAAHCFMTNGTLEPAEWSVLGVHSDQDPLDGA 117
Db 56 WPMQVSIITGVHVCGLSVNQVWVSAHCFPREHSE---EYEVKLGHQHDSYNDI 112
Qy 118 HTRAVAAIVVPANYSQVELGADLALLRLASPAISGPAVWPVCLPRASHRFVHGTCWATG 177
Db 113 VVHTVAQIITHSSYREEGSQDIALIRLSBPVTSRVRIRICLPAANASFPNGLHCTVTG 172
Qy 178 WGDVQEAADPLPLPWVLQVEVRLILGEATCOCLYSQPGFNLTLQILPGMLCAGYPEGRRD 237
Db 173 WGHVAPSVSLQTPRPLQOQLEVPILSRFCSCLYNINAVPEEPHTIQDMLCAGYVKGKD 232
Qy 238 TCQDGGGPLYCEEGGRWFQAGITSFGCGRRNRPGVFTAVATYEAWIREQVMSBPFP 297
Db 233 ACQDGGGPLYCEEGGRWFQAGITSFGCGRRNRPGVFTAVATYEAWIREQVMSBPFP 297
Qy 298 AFPTQPKTQSD---CLHQTAF-LDSARILLRPLSHISVGVSTGKSLVLPWL 346
Db 291 RVVPQTQESQPDGHLGNHHPVFSSAAAPKLLRPVLFPLGLTGLLSL---WL 340

RESULT 8
US-10-051-874-87
; Sequence 87, Application US/10051874
; Publication No. US20040005557A1
; GENERAL INFORMATION:
; APPLICANT: Padigaru, Muralidhara
; APPLICANT: Alsobrook II, John P
; APPLICANT: Colman, Steven D
; APPLICANT: Spytek, Kimberly A
; APPLICANT: Boldog, Ferenc
; APPLICANT: Vernet, Corine AM
; APPLICANT: Li, Li
; APPLICANT: Shency, Suresh G
; APPLICANT: Casman, Stacie J
; APPLICANT: Guo, Xiaojia Sasha
; APPLICANT: Edinger, Shlomit R
; APPLICANT: MacDougall, John R
; APPLICANT: Malyankar, Uriel M
; APPLICANT: Patturajan, Meera
; APPLICANT: Shimkets, Richard A
; APPLICANT: Pena, Carol EA
; APPLICANT: Tchernev, Velizar T
; APPLICANT: Zerhusen, Bryan D

LENGTH: 34Z
TYPE: PRF
APPLICANT: Burgess, Catherine E
; APPLICANT: Burgess, Catherine E
INTERVIEW OF INTERVIEWING AGENCIES ENCLOSURE ENCODING THEM AND METHODS OF

Db 205 QAERI---VLPGLCAGYPQGHKDACQSDSGGLTCLQSGSWLVGVVSWGKGCALPNRP 261
QY 274 GVFTAVATYEAWAREV 290
Db 262 GYVTSVATYSPWIOARV 278

RESULT 11
US-10-041-264A-7
; Sequence 7, Application US/10041264A
; Publication No. US20020142446A1
; GENERAL INFORMATION:
; APPLICANT: Darrow, Andrew
; APPLICANT: Andrade-Gordon, Patricia
; APPLICANT: Qi, Jensen
; TITLE OF INVENTION: DNA Encoding the Human Serine
; FILE REFERENCE: ORT-1031
; CURRENT APPLICATION NUMBER: US/10/041,264A
; CURRENT FILING DATE: 2002-01-08
; PRIOR APPLICATION NUMBER: US/09/387,375
; PRIOR FILING DATE: 1999-08-31
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 7
; LENGTH: 284
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-10-041-264A-7

Query Match 30.2%; Score 590; DB 13; Length 284;
Best Local Similarity 46.3%; Pred. No. 4.3e-43;
Matches 119; Conservative 29; Mismatches 99; Indels 10; Gaps 3;

QY 38 CGPEPSARIYGGSNAGPQGTWPMQVSLHHGGHICGSLIAPSWVLAAHCFMTNGTLEP 97
Db 28 CGQPRMSRIYGGDRDGEWPMQASIQHPGAHVCGGSLIAPQWVLTAAHCFPRAL--- 84

QY 98 AAWSVLLGVHSGDGLDGAHTRAVAAIVVPANYSOVELGADLALLRLASPASIGPAVWP 157
Db 85 PAEYRVLGALRGSTSPRTLSVFRVLLPPDYSEDGARGDLALLQLRRPVLSARVQP 144

QY 158 VCLPRASHRFVHGATWAGDVQADPLPLPWVQLVEVLRLIGATCCQCLY----SQP 213
Db 145 VCLPVGARPPPGTPCRVTGWSLRPGVLPPEWRPLOGVRVPLDSTRCDGLYHVGADVP 204

QY 214 GPFNLTLQILPGLMCAGYPEGRRTDCCQSDSGPLVCEEGGWFOAGITSFGGCGRRNRP 273
Db 205 QAERI---VLPGLCAGYPQGHKDACQSDSGGLTCLQSGSWLVGVVSWGKGCALPNRP 261

QY 274 GVFTAVATYEAWAREV 290
Db 262 GYVTSVATYSPWIOARV 278

RESULT 12
US-10-042-091A-7
; Sequence 7, Application US/10042091A
; Publication No. US20020142447A1
; GENERAL INFORMATION:
; APPLICANT: Darrow, Andrew
; APPLICANT: Andrade-Gordon, Patricia
; APPLICANT: Qi, Jensen
; TITLE OF INVENTION: DNA Encoding the Human Serine
; FILE REFERENCE: ORT-1031
; CURRENT APPLICATION NUMBER: US/10/042,091A
; CURRENT FILING DATE: 2002-01-08
; PRIOR APPLICATION NUMBER: US/09/387,375
; PRIOR FILING DATE: 1999-08-31
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: PatentIn Ver. 2.0

QY 38 CGPEPSARIYGGSNAGPQGTWPMQVSLHHGGHICGSLIAPSWVLAAHCFMTNGTLEP 97
Db 28 CGQPRMSRIYGGDRDGEWPMQASIQHPGAHVCGGSLIAPQWVLTAAHCFPRAL--- 84

QY 98 AAWSVLLGVHSGDGLDGAHTRAVAAIVVPANYSOVELGADLALLRLASPASIGPAVWP 157
Db 85 PAEYRVLGALRGSTSPRTLSVFRVLLPPDYSEDGARGDLALLQLRRPVLSARVQP 144

QY 158 VCLPRASHRFVHGATWAGDVQADPLPLPWVQLVEVLRLIGATCCQCLY----SQP 213
Db 145 VCLPVGARPPPGTPCRVTGWSLRPGVLPPEWRPLOGVRVPLDSTRCDGLYHVGADVP 204

QY 214 GPFNLTLQILPGLMCAGYPEGRRTDCCQSDSGPLVCEEGGWFOAGITSFGGCGRRNRP 273
Db 205 QAERI---VLPGLCAGYPQGHKDACQSDSGGLTCLQSGSWLVGVVSWGKGCALPNRP 261

QY 274 GVFTAVATYEAWAREV 290
Db 262 GYVTSVATYSPWIOARV 278

RESULT 10
US-10-041-400A-7
; Sequence 7, Application US/10041400A
; Publication No. US20020110895A1
; GENERAL INFORMATION:
; APPLICANT: Darrow, Andrew
; APPLICANT: Andrade-Gordon, Patricia
; APPLICANT: Qi, Jensen
; TITLE OF INVENTION: DNA Encoding the Human Serine
; FILE REFERENCE: ORT-1031
; CURRENT APPLICATION NUMBER: US/10/041,400A
; CURRENT FILING DATE: 2002-01-08
; PRIOR APPLICATION NUMBER: US/09/387,375
; PRIOR FILING DATE: 1999-08-31
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 7
; LENGTH: 284
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-10-041-400A-7

Query Match 30.2%; Score 590; DB 13; Length 284;
Best Local Similarity 46.3%; Pred. No. 4.3e-43;
Matches 119; Conservative 29; Mismatches 99; Indels 10; Gaps 3;

QY 38 CGPEPSARIYGGSNAGPQGTWPMQVSLHHGGHICGSLIAPSWVLAAHCFMTNGTLEP 97
Db 28 CGQPRMSRIYGGDRDGEWPMQASIQHPGAHVCGGSLIAPQWVLTAAHCFPRAL--- 84

QY 98 AAWSVLLGVHSGDGLDGAHTRAVAAIVVPANYSOVELGADLALLRLASPASIGPAVWP 157
Db 85 PAEYRVLGALRGSTSPRTLSVFRVLLPPDYSEDGARGDLALLQLRRPVLSARVQP 144

QY 158 VCLPRASHRFVHGATWAGDVQADPLPLPWVQLVEVLRLIGATCCQCLY----SQP 213
Db 145 VCLPVGARPPPGTPCRVTGWSLRPGVLPPEWRPLOGVRVPLDSTRCDGLYHVGADVP 204

QY 214 GPFNLTLQILPGLMCAGYPEGRRTDCCQSDSGPLVCEEGGWFOAGITSFGGCGRRNRP 273

```

; SEQ ID NO 7
; LENGTH: 284
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-10-042-091A-7

Query Match
Best Local Similarity 30.2%; Score 590; DB 13; Length 284;
Matches 119; Conservative 29; Mismatches 99; Indels 10; Gaps 3;

QY 38 CGRPEPARIIVGSSNAQPTWQVSLHGGGHCIGSLIAPSWLSAAHCFMTNGTLEP 97
Db 28 CGQPMSSRIIVGRDGDGSEWPMQASIQHPGAHVGCSLIAPQWVLTAAHCFPRAL--- 84

QY 98 AAEWSVLGVHSGDGLDGHATRAVAIVVPANYSOVELGADLALLRLASPSLIGPAVWP 157
Db 85 PAEYRVLGALRLGSTSPRTLSVFEVRVLLPPDYSEDGARGDLALLQLRRPVLARVQP 144

QY 158 VCLPRASHRFVHGTACWATGWDVQADPLPLFWVLQEVRLRLGATCCCLY---SQP 213
Db 145 VCLPVPGARPPPTGTCRVTCWGLSLPGVPLPEWRPLQGVVRVPLDGRTCGLYHVGADVP 204

QY 214 GPNLTLQLPGLMCLAGYPERDTCQDSDGGLVCEEGRWFCAGITSEFGGGRNRP 273
Db 205 QAERI---VLPGLSLCAGYPQGHKDACQDSDGGLTCLQSGSVLGVVSWGKCALPNRP 261

QY 274 GVFTAVATYEAWIREQV 290
Db 262 GVTISVATYSPWIQARV 278

```

```

RESULT 13
US-10-051-874-89
; Sequence 89, Application US/10051874
; Publication No. US2004000557A1
; GENERAL INFORMATION:
; APPLICANT: Padigaru, Muralidhara
; APPLICANT: Alsobrook II, John P
; APPLICANT: Colman, Steven D
; APPLICANT: Spytsek, Kimberly A
; APPLICANT: Boldog, Ferenc
; APPLICANT: Vernet, Corine AM
; APPLICANT: Li, Li
; APPLICANT: Shenoy, Suresh G
; APPLICANT: Casman, Stacie J
; APPLICANT: Guo, Xiaohua Saeba
; APPLICANT: Edinger, Shlomit R
; APPLICANT: MacDougall, John R
; APPLICANT: Malyankar, Uriel M
; APPLICANT: Patturajan, Meera
; APPLICANT: Shimkets, Richard A
; APPLICANT: Pena, Carol EA
; APPLICANT: Tchernev, Velizar T
; APPLICANT: Zerhusen, Bryan D
; APPLICANT: Millet, Isabelle
; APPLICANT: Miller, Charles E
; APPLICANT: Lepley, Denise M
; APPLICANT: Smithson, Glenda
; APPLICANT: Baumgartner, Jason C
; APPLICANT: Herrman, John L
; APPLICANT: Peyman, John A
; APPLICANT: Gorman, Linda
; APPLICANT: Mezes, Peter D
; APPLICANT: Kekuda, Ramesh
; APPLICANT: Taupier Jr, Raymond J
; APPLICANT: Gerlach, Valerie
; APPLICANT: Grosse, William M
; APPLICANT: Liu, Xiaohong
; APPLICANT: Ellerman, Karen
; APPLICANT: Rothenberg, Mark
; APPLICANT: Stone, David J
; APPLICANT: Burgess, Catherine E
; TITLE OF INVENTION: PROTEINS, POLYNUCLEOTIDES ENCODING THEM AND METHODS OF

```

```

; TITLE OF INVENTION: USING THE SAME
; FILE REFERENCE: 21402-245
; CURRENT APPLICATION NUMBER: US/10/051,874
; CURRENT FILING DATE: 2002-09-25
; PRIOR APPLICATION NUMBER: 60/268,595
; PRIOR FILING DATE: 2001-02-14
; PRIOR APPLICATION NUMBER: 60/325,306
; PRIOR FILING DATE: 2001-09-27
; PRIOR APPLICATION NUMBER: 60/262,587
; PRIOR FILING DATE: 2001-01-18
; PRIOR APPLICATION NUMBER: 60/272,409
; PRIOR FILING DATE: 2001-02-28
; PRIOR APPLICATION NUMBER: 60/262,454
; PRIOR FILING DATE: 2001-01-18
; PRIOR APPLICATION NUMBER: 60/276,777
; PRIOR FILING DATE: 2001-03-16
; PRIOR APPLICATION NUMBER: 60/291,672
; PRIOR FILING DATE: 2001-05-17
; PRIOR APPLICATION NUMBER: 60/330,336
; PRIOR FILING DATE: 2001-10-18
; PRIOR APPLICATION NUMBER: 60/265,530
; PRIOR FILING DATE: 2001-01-31
; PRIOR APPLICATION NUMBER: 60/261,376
; PRIOR FILING DATE: 2001-01-16
; NUMBER OF SEQ ID NOS: 269
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 89
; LENGTH: 285
; TYPE: PRT
; ORGANISM: Mus musculus
; US-10-051-874-89

Query Match 29.9%; Score 583.5; DB 15; Length 285;
Best Local Similarity 43.3%; Pred. No. 1.6e-42;
Matches 127; Conservative 39; Mismatches 116; Indels 11; Gaps 6;

QY 1 MAOKGVLPQGLGAVANSYSILYLPVG-PARG-PPYCGRP-EPSARVGGSNAPGT 57
Db 1 MALRVGLGLQLEAVT---ILLGLLQSGIRADGTASCQAVIQP--RITGGGSAKPGQ 55

QY 58 WPMQVSLHGGGHCIGSLIAPSWLSAAHCFMTNGTLEPAEWSVLGVHSGDGLDGA 117
Db 56 WPMQVSLHGGGHCIGSLIAPSWLSAAHCFMTNGTLEPAEWSVLGVHSGDGLDGA 112

QY 118 HTRAVAIIVVPANYSOVELGADLALLRLASPSLIGPAVWPVCLPRASHRFVHGTACWATG 177
Db 113 VVHTVAQIITHSSVREBSQGDIALIRLSFVTFSYRIPICLPAANASFPNGLHCTVTG 172

QY 178 WGVQVQADPLPLFWVLQEVRLRLGATCCCLYSDQPGFNLTLQLPGLMCLAGYPERD 237
Db 173 WGVQVQADPLPLFWVLQEVRLRLGATCCCLYSDQPGFNLTLQLPGLMCLAGYPERD 232

QY 238 TCQDSDGGLVCEEGRWFCAGITSEFGGGRNRPQVFTAVATYEAWIREQV 290
Db 233 ACQDSDGGLVCEEGRWFCAGITSEFGGGRNRPQVFTAVATYEAWIREQV 285

```

```

RESULT 14
US-10-311-035-8
; Sequence 8, Application US/10311035
; Publication No. US2004002243A1
; GENERAL INFORMATION:
; APPLICANT: INCYTE GENOMICS, INC.
; APPLICANT: YUE, Henry
; APPLICANT: ELLIOTT, Vicki
; APPLICANT: GANDHI, Ameena R.
; APPLICANT: LAL, Preeti
; APPLICANT: AU-YOUNG, Janice
; APPLICANT: TRIBOULEY, Catherine M.
; APPLICANT: DELEGEANE, Angelo M.
; APPLICANT: BAUGHN, Mariah R.
; APPLICANT: NGUYEN, Darniel B.
; APPLICANT: LEE, Ernestine A.

```

Thu Feb 26 13:50:28 2004

```

; APPLICANT: HAFALIA, April
; APPLICANT: KHAN, Farrah A.
; APPLICANT: CHAWLA, Narinder K.
; APPLICANT: YAO, Monique G.
; APPLICANT: LU, Dyung Aina M.
; APPLICANT: ARVIZU, Chandra S.
; APPLICANT: TANG, Y. Tom
; APPLICANT: WALSH, Roderick T.
; APPLICANT: AZIMZAI, Yalda
; APPLICANT: LU, Yan
; APPLICANT: RAMKUMAR, Jayalaximi
; APPLICANT: XU, Yuming
; APPLICANT: REDDY, Roopa
; APPLICANT: DAS, Depodriya
; APPLICANT: KEARNEY, Liam
; APPLICANT: KALLICK, Deborah A.
; TITLE OF INVENTION: Proteases
; FILE REFERENCE: PI-0123 PCT
; CURRENT APPLICATION NUMBER: US/10/311,035
; CURRENT FILING DATE: 2002-12-10
; PRIOR APPLICATION NUMBER: 60/212,336; 60/213,955; 60/215,396; 60/216,821; 60/218,946
; PRIOR FILING DATE: 2000-06-16; 2000-06-22; 2000-06-29; 2000-07-07; 2000-07-14
; NUMBER OF SEQ ID NOS: 42
; SOFTWARE: PERL Program
; SEQ ID NO 8
; LENGTH: 346
; TYPE: PRT
; ORGANISM: Homo sapiens
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No. US2004002343A1 2256251CD1
US-10-311-035-8

Query Match      29.8%; Score 582.5; DB 16; Length 346;
Best Local Similarity 43.8%; Pred. No. 2.4e-42;
Matches 133; Conservative 38; Mismatches 104; Indels 29; Gaps 12;

QY 10 GQLGAVANDSVLSYGLVSPGARGPPYCGREPS---ARIIVGSSNAQPGTWPQVSLHH 66
DB 24 GQ-GRILHGSVAVGLSPGPTGSSFDLGCGRPOVSDAGRIVGGHAAPAGAPWQASRL 82
QY 67 GGCHICGSLIAPSVLSAAHCFMTNGTLEPAAEWSVLLGVHSDQGLD---GAHTRAVA 123
DB 83 RRVHVGCSLLSPQWLTAAHCF--SSLN-SSDYQVHL-----GELEITLSPHSTVR 133
QY 124 AIIVVPANYS--QVELGADLALLRLASPASLGPAPVWVCLPRASHRFVHGTAACWATGCDVQ 182
DB 134 QIILHSPSPGPGTSGDIALVELSVPTLFRILPVLCPPEASDDFCFGIRCWVTGWYTR 193
QY 183 EADPLPLPWLQVEVELLGEATCCLYSOPGPNLTLOILPGMLCAGYPEGRRDTCQGD 242
DB 194 EGBELPPPYSLREVKVSVVDTECRRDY--PGPGGSLIQ--PMLCA--RFGDACQDD 246
QY 243 SGGPLVCEGGRMPQAGITSFSGCGRRNRPGVFTAVATYEAWIREQVM---GSEPGPAF 299
DB 247 SGGPLVQVNGAWYQAGIVSWGEGCGRRNRPVGVYTRPAYVNVIRRHITASGGSESG--Y 304
QY 300 PTQP 303
DB 305 PRLP 308

RESULT 15
US-10-221-097-36
; Sequence 36, Application US/10221097
; Publication No. US20030144476A1
; GENERAL INFORMATION:
; APPLICANT: Agarwal, Pankaj
; APPLICANT: Murdock, Paul R.
; APPLICANT: Rizvi, Safia K.
; APPLICANT: Smith, Randall F.
; APPLICANT: Xiang, Zhaoqing
; TITLE OF INVENTION: NOVEL COMPOUNDS

```

```

; FILE REFERENCE: GP50016
; CURRENT APPLICATION NUMBER: US/10/221,097
; CURRENT FILING DATE: 2002-09-06
; PRIOR APPLICATION NUMBER: PCT/US01/07143
; PRIOR FILING DATE: 2001-03-05
; PRIOR APPLICATION NUMBER: 60/187,107
; PRIOR FILING DATE: 2000-03-06
; PRIOR APPLICATION NUMBER: 60/236,874
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/188,916
; PRIOR FILING DATE: 2000-03-13
; PRIOR APPLICATION NUMBER: 60/237,846
; PRIOR FILING DATE: 2000-10-03
; NUMBER OF SEQ ID NOS: 52
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 36
; LENGTH: 255
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-10-221-097-36

Query Match      29.5%; Score 575.5; DB 14; Length 255;
Best Local Similarity 43.7%; Pred. No. 6.8e-42;
Matches 128; Conservative 26; Mismatches 88; Indels 51; Gaps 9;

QY 47 IVGSSNAQPGTWPQVSLHGGGHCIGGSLIAPSVLSAAHCFMTNGTLEPAAEWSVLLG 106
DB 1 IVGSSNAQPGTWPQVSLHGGGHCIGGSLIAPSVLSAAHCFMTG-----46
QY 107 VHSQDGLDGAHTRAVAIVVPAN-----YSQVELGADLALLRLASPASLGPAPVWVCLP 161
DB 47 --RQVRCPEPTRRTRS-----ALPTRKRRRAYNHYSQSDALLQLAHPT---HTPLCLP 96
QY 162 RASHRFVHGTAACWATGWC-DVQEADPL-PLPWLQVEVELLGEATCCLYSQPGPNLT 219
DB 97 QPAHFFPGASCWATGWDQDTSAPSLSPAGTLENLRLUISRPTCNCIYNQLHORHLS 156
QY 220 LQILFGLMCAGYPEGRRDTCQDSDGGLVVCBEGGWFOAGITSFSGCGRRNRPGVFTAV 279
DB 157 NPAPFGMLCGGPGQPGVQCGFLGAPLVHEVRGTWFLAGLHSGDACCQGPAPPAVFTAL 216
QY 280 ATYE-----AWIREQVMSEPGPAFTQPKTQSDCLHQTAFDLSARILLRPLS 328
DB 217 PAMRTGSAVWTRQVYFABEP-----EPEAEFGSCLAN-----IRPFS 253

Search completed: February 25, 2004, 16:08:53
Job time : 642 secs

```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: February 25, 2004, 15:51:34 ; Search time 3779 Seconds
(without alignments)
4094.592 Million cell updates/sec

Title: US-10-037-417-46

Perfect score: 1953

Sequence: 1 MAQKVLGPGQLGAVANSDS.....TKSLVLPWLSPHSLGLWGF 357

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 3470272 seqs, 21671516995 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlbp
-Q=/cgn2.1/USPTO.spool.p/US10037417/runat_24022004_101038_17447/app_query.fasta_1.519
-DB=GenEmbl -QFMT=fastap -SUFFIX=p2n.rge -MINMATCH=0.1 -LOOPEL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptco -NORM=ext -HEARSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10037417@cgn_1_1_4958 @runat_24022004_101038_17447 -NCPU=6 -ICPU=3
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

GenEmbl.*

1: gb_ba.*

2: gb_htg.*

3: gb_in.*

4: gb_om.*

5: gb_ov.*

6: gb_pat.*

7: gb_ph.*

8: gb_pl.*

9: gb_pr.*

10: gb_ro.*

11: gb_sts.*

12: gb_sy.*

13: gb_un.*

14: gb_vi.*

15: em_ba.*

16: em_fun.*

17: em_hum.*

18: em_in.*

19: em_mu.*

20: em_om.*

21: em_or.*

22: em_ov.*

23: em_pat.*

24: em_ph.*

25: em_pl.*

26: em_ro.*

27: em_sts.*

28: em_un.*

29: em_vi.*
30: em_htg_hum.*
31: em_htg_inv.*
32: em_htg_other.*
33: em_htg_mus.*
34: em_htg_pln.*
35: em_htg_rdi.*
36: em_htg_mam.*
37: em_htg_vrt.*
38: em_sy.*
39: em_htgo_hum.*
40: em_htgo_mus.*
41: em_htgo_other.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query %	Match	Length	DB	ID	Description
1	1953	100.0	1102	6	AX512287	AX512287	Sequence
2	1953	100.0	1102	6	AX512289	AX512289	Sequence
3	1505	77.1	2457	6	AX360096	AX360096	Sequence
4	1505	77.1	2681	6	AX480935	AX480935	Sequence
5	1492	76.4	2810	6	BD127529	BD127529	Primer for
6	1492	76.4	2810	9	AK075142	AK075142	Homo sapiens
C 7	970	49.7	537	6	AX342934	AX342934	Sequence
8	861	44.1	670	6	BD125219	BD125219	Primer for
9	861	44.1	670	6	BD126485	BD126485	Primer for
C 10	799	40.9	456	6	AX342936	AX342936	Sequence
C 11	716.5	36.7	127769	9	AC009088	AC009088	Homo sapiens
12	697	35.7	1796	6	AX098193	AX098193	Sequence
13	697	35.7	1800	9	BC001462	BC001462	Homo sapiens
14	697	35.7	1834	6	AX335777	AX335777	Sequence
15	697	35.7	1834	6	AX336076	AX336076	Sequence
16	697	35.7	1834	6	AX336340	AX336340	Sequence
17	697	35.7	1834	6	AX409567	AX409567	Sequence
18	697	35.7	1834	6	AX474697	AX474697	Sequence
19	697	35.7	1834	9	HUMPROS	I41351	Homo sapiens
20	697	35.7	1835	6	AX098215	AX098215	Sequence
C 21	658	33.7	232119	2	AC106629	AC106629	Rattus norvegicus
22	652	33.4	1726	6	AX675579	AX675579	Sequence
23	630.5	32.3	1727	10	AY335911	AY335911	Mus musculus
24	630.5	32.3	1743	10	BC003851	BC003851	Mus musculus
25	630.5	32.3	1856	10	AB038244	AB038244	Mus musculus
26	630.5	32.3	1870	10	AF378085	AF378085	Mus musculus
27	627.5	32.1	1797	10	AF188613	AF188613	Mus musculus
28	621	31.8	1864	10	AF202076	AF202076	Rattus norvegicus
29	621	31.8	1892	10	BC061800	BC061800	Rattus norvegicus
C 30	620.5	31.8	185788	10	AC124461	AC124461	Mus musculus
31	620.5	31.8	219763	10	AC093175	AC093175	Mus musculus
32	620	31.7	2208	10	AB017638	AB017638	Rattus norvegicus
33	613	31.4	1445	10	AY266139	AY266139	Mus musculus
34	612	31.3	1161	6	AX675581	AX675581	Sequence
35	611	31.3	244161	2	AC117170	AC117170	Rattus norvegicus
36	609	31.2	1360	10	AY262280	AY262280	Mus musculus
37	608	31.1	1488	10	BC024903	BC024903	Mus musculus
38	603.5	30.9	1606	6	AX676264	AX676264	Sequence
39	603.5	30.9	1613	6	AR256990	AR256990	Sequence
40	603.5	30.9	1613	9	AF536382	AF536382	Homo sapiens
41	603.5	30.9	1718	9	BC062334	BC062334	Homo sapiens
42	603.5	30.9	1755	9	BC036846	BC036846	Homo sapiens
43	597	30.6	944	6	AX375744	AX375744	Sequence
44	591	30.3	1020	6	AX823252	AX823252	Sequence
45	590.5	30.2	1458	5	BC056000	BC056000	Xenopus laevis

ALIGNMENTS

RESULT 1

US-10-037-417-46 (1-357) x AX512289 (1-1102)

QY 1 MetalGlnLysGlyValLeuGlyProGlyGlnLeuGlyAlaValAlaAsnSerAspSer 20

Db 19 ATGGCCCAAGAGGGGGTCTTGGGGCTGGGCAGCTGGGGCTGTGCCAATTCTGACTCA 78

QY	258	AlaGlyTleThrSerPheGlyPheGlyCysGlyArgArgAsnArgProGlyValPheThr	277
Db	772	GCAGGANTCACCAGCTTTGGCTTTGGCTGGAGCGGAGAAACCGCCCTGGAGTTTCACT	831
QY	278	AlaValAlaThrTyrGluAlaTrrPileArgGluGlnValMetGlySerGluProGlyPro	297
Db	832	GCTGGCTACTATGAGGATGATGATGAGGATGATGAGGATGATGAGGATGATGAGGAT	891
QY	298	AlaPheProThrGlnProGlnLysThrGlnSerAsp	309
Db	892	GCCTTTCCACCCAGCCCGCCAGAGACCCAGTCAGAT	927
RESULT 4			
LOCUS	AX480935	2681 bp	DNA linear PAT 12-AUG-2002
DEFINITION	Sequence 31 from Patent WO0246383.		
ACCESSION	AX480935		
VERSION	AX480935.1	GI:22217574	
KEYWORDS	Homo sapiens (human)		
SOURCE	Homo sapiens		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
AUTHORS	Yue, H., Azimzai, Y., Kallick, D.A., Baughn, M.R., Griffin, J.A., Swarnakar, A., Lai, P.G., Walla, N.K., Hafalia, A.J., Gandhi, A.R., Au-Young, J., Elliott, V.S., Ramkumar, J., Thangavelu, K., Lu, Y., Warren, B.A., Lu, D.A., Lee, E.A., Tribouley, C.M., Arvizu, C., Deleane, A.M., Yao, M.G., Khan, F.A. and Sanjanwala, M.M.		
TITLE	Protein modification and maintenance molecules		
JOURNAL	Patent: WO 0246383-A 31 13-JUN-2002; Incyte Genomics, Inc. (US)		
FEATURES	Location/Qualifiers		
source	1..2681		
ORGANISM	/organism="Homo sapiens"		
mol_type	/mol_type="unassigned DNA"		
db_xref	/db_xref="taxon:9606"		
note	/note="Incyte ID No: 2751509CB1"		
ORIGIN			
Alignment Scores:	5.72e-73	Length:	2681
Pred. No.:	1505.00	Matches:	272
Score:	100.00%	Conservative:	0
Percent Similarity:	100.00%	Mismatches:	0
Best Local Similarity:	100.00%	Indels:	0
Query Match:	77.06%	Gaps:	0
DB:	6		
US-10-037-417-46 (1-357) x AX480935 (1-2681)			
QY	38	CysGlyArgProGluProSerAlaArgileValGlyGlySerAsnAlaGlnProGlyThr	57
Db	112	TGCGGGCGCCCTGAGCCCTCGGCCCGCATCGTGGGGGGCTCAAACGCGCAGCGGGCACC	171
QY	58	TrpProThrGlnValSerLeuHisGlyGlyGlyHisGlyGlyGlyGlyGlyGlyGlyGly	77
Db	172	TGGCGCTTGGCAAGTGAGCGCTGCACATGAGGAGTGCCACATCTCGGGGGCTCCCTCATC	231
QY	78	AlaProSerTrpValLeuSerAlaAlaHisCysPheMetThrAsnGlyThrLeuGluPro	97
Db	232	GCCCGCTCTGGGCTCTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT	291
QY	98	AlaAlaGluTrpSerValLeuGlyValHisSerGlnAspGlyProLeuAspGlyAla	117
Db	292	GCGGCCGAGTGGTGGTACTGCTGGGCGTGACATCCAGGACGGGGCCCTCGAGCGCGC	351
QY	118	HisThrArgAlaValAlaAlaValValProAlaAsnTyrSerGlnValGluLeuGly	137
Db	352	CACACCCGCGAGTGGCGGCCCTCGTGGTGGCGGCCAATACAGCAAGTGGAGCTGGGC	411
QY	138	AlaAspLeuAlaLeuLeuArgLeuAlaSerProAlaSerLeuGlyProAlaValTrpPro	157

ACCESSION AK075142
 VERSION AK075142.1 GI:22761040
 KEYWORDS oligo capping; fis (full insert sequence).
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1
 AUTHORS Isogai,T., Ota,T., Nishikawa,T., Hayashi,K., Otsuki,T., Sugiyama,T., Suzuki,Y., Nagai,K., Sugano,S., Ishii,S., Kawai-Hio,Y., Saito,K., Yamamoto,J., Wakamatsu,A., Nakamura,Y., Kojima,S., Nagahari,K., Masuho,Y., Ono,T., Okano,K., Yoshikawa,Y., Aotaka,S., Sasaki,N., Hattori,A., Okumura,K., Iwayanagi,T. and Ninomiya,K.
 NEDO human cDNA sequencing project
 Unpublished
 JOURNAL
 REFERENCE
 AUTHORS Isogai,T. and Otsuki,T.
 TITLE Direct Submission
 JOURNAL Submitted (25-MAR-2002) Takao Isogai, Helix Research Institute, Genomics Laboratory; 1532-3 Yana, Kisarazu, Chiba 292-0812, Japan (E-mail:genomics@hri.co.jp, tel:81-438-52-3975, fax:81-438-52-3986)
 COMMENT NEDO human cDNA sequencing project supported by Ministry of Economy, Trade and Industry of Japan; cDNA full insert sequencing: Research Association for Biotechnology; cDNA library construction: Institute of Medical Science, University of Tokyo, Laboratory of Genome Structure, Human Genome Center; cDNA 5'- & 3'-end one pass sequencing and clone selection: Helix Research Institute (supported by Japan Key Technology Center etc.).

FEATURES
 Location/Qualifiers
 1..2810
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="PLACE1005003"
 /tissue type="placenta"
 /clone_lib="PLACE1"
 /note="Cloning vector: pME18SFL3"
 300..2600
 /note="unnamed protein product"
 /codon_start=1
 /protein_id="BAC11431.1"
 /db_xref="GI:22761041"
 /translations="MTNGTLEPAAEWSVLGVHSDGDLPGARHVAALVVPANTSGV
 VELGADALLRLASPAAGVAVPVCLPRASHRFVHGTCATWGTGVDQADPLPLFV
 VLQVEURLLGEATCCQLSQPGPNLTQLILPGMLCAGYPEGRRDTCCGDSGGPLVC
 EGGRRFQAGITSGFGCGRRNRPQVFTAVATYEAIRQVMSGEPGPAFPTQPKTO
 SDQPEARENCTIALPECGARRPCHGALVQVWVPGSRPCHGALVSESWVLAPASCF
 LDPSNDSPPRDLDAWVLLSPRAERAVRQVWVPGSRPCHGALVSESWVLAPASCF
 AASRPVCLPPEHVFILPGRSCRLARWGRGEPALGAEILAEILGCMWCHGLRPPRA
 AVPLPGDPHALCPAYQEEKVSQWNSRWLLCOBEGTWFLAGIDFGGCLGLRPPRA
 FFPLQTHGPMWISHTRGAYLEQLDQALNDWPGDGETETQCTPHTHEGACGLLEAAV
 GVLMPWLAEVHADRVCTGILLAGPWLAATHCVLBPGSTTVPYIEVILGRAGASSL
 PQGQVSRVLSIRLPOHLGRLPALLELSRVPESPAIPICLHPAGIPPGASCVW
 LGWKEPDQRPVFAAAVSLTQRCDCLYQGLPPGLTCLVLYAEGQENRCENTWAPPLI
 COMTEGSSWLVGMVQGSRELFAIGPEAWISQTVGEANLPPSGSPHWPMTGGSLUC
 PPELAKAGSPHAYVFLLLLTLLIQS"

ORIGIN
 Alignment Scores:
 Pred. No.: 3,02e-72 Length: 2810
 Score: 1492.00 Matches: 272
 Percent Similarity: 99.63% Conservative: 0
 Best Local Similarity: 99.63% Mismatches: 0
 Query Match: 76.40% Indels: 1
 DB: 9 Gaps: 0

US-10-037-417-46 (1-357) x AK075142 (1-2810)

QY 38 CysGlyArgProGluProSerAlaAlaGlyVal-GlyGlySerAsnAlaGlnProGlyTh 57
 Db 143 TCGGGCGCCCTGAGCCCTCGGCCGCGCATCGTGGGGGGGCTCAACGCGCAGCGCGGCAC 202

QY 57 rTrpProTrpGlnValSerLeuHisHisGlyGlyGlyHisileCysGlySerLeu1 77
 Db 203 CTGGCCCTTGGCAGTGGAGCTGCACCATGAGGTGGCCCATCTCGCGGGGGCTCCCTCAT 262

QY 77 eAlaProSerTrpValLeuSerAlaAlaHisCysPheMetThrAsnGlyThrLeuGluPr 97
 Db 263 CGCCCCCTCTCGGGTCTCTCCGCTGCTCACTGTTTCATGCAATGGAGCGCTGGAGCC 322

QY 97 oAlaAlaGluTrpSerValLeuLeuGlyValHisSerGlnAspGlyProLeuAspGlyAl 117
 Db 323 CGCGCCGAGTGGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 382

QY 117 aHisThrArgAlaValAlaAlaValValProAlaAsnTyrSerGlnValGluLeuG1 137
 Db 383 GCACACCCGCGAGTGGCGCCCATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 442

QY 137 yAlaAspLeuAlaLeuLeuArgLeuAlaSerProAlaSerLeuGlyProAlaValTrpPr 157
 Db 443 CGCCGACTTGGCCCTTGGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 502

QY 157 oValCysLeuProArgAlaSerHisArgPheValHisGlyThrAlaCysTrpAlaThrGl 177
 Db 503 TGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 562

QY 177 yTrpGlyAspValGlnGluAlaAspProLeuProTrpValLeuGlnGluValGl 197
 Db 563 CTGGGAGACGCTCCAGGAGGAGATCTCTGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 622

QY 197 nLeuArgLeuLeuGlyGluAlaThrCysGlnCysLeuTrpSerGlnProGlyProPheAs 217
 Db 623 GTAAAGGCTGCTGGGGGAGGCCCTCTGCTCAATGCTCTACAGCCAGCCCGCTCCCTCAA 682

QY 217 nLeuThrLeuGlnLeuLeuProGlyMetLeuCysAlaGlyTyrProGluGlyArgArgAs 237
 Db 683 CTCTACTCTCCAGATATTGCCAGGATGCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 742

QY 237 pThrCysGlnGlyAspSerGlyGlyProLeuValCysGlnGluGlyGlyArgTrpPheGl 257
 Db 743 CACTCTGCCAGGCGCATCTCTGGGGGGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 802

QY 257 nAlaGlyIleThrSerPheGlyPheGlyCysGlyArgArgAsnArgProGlyValPheTh 277
 Db 803 GGCAGGAATCACCATCTTTGGCTTTTGGCTGTGGACGAGAACCCGCTGGAGTTTTCAC 862

QY 277 xAlaValAlaThrTyrGluAlaTrpIleArgGluGlnValMetGlySerGluProGlyPr 297
 Db 863 TGCTGTGGCTACCTATCAGGCATGATACGGAGCAGGTGATGGTTTCAGAGCCTGGGCC 922

QY 297 oAlaPheProThrGlnProGlnLysThrGlnSerAsp 309
 Db 923 TGCCTTTCCACCGACGCGCCCGAGAGACCCAGTCAGAT 959

RESULT 6
 AK075142 2810 bp mRNA linear PRI 03-SEP-2002
 LOCUS AK075142
 DEFINITION Homo sapiens cDNA FLJ30661 fis, clone PLACE1005003, weakly similar to PROSTASIN PRECURSOR (EC 3.4.21.-).

ORIGIN

203	CTGGCCCTTGGCAAGTGAAGCCTTGACCATGAGGTGGCCACATCTGCGGGGGCTCCCTCAT	262
77	eAlaProSerTrpValLeuSerAlaAlaHisCysPheMetThrAsnGlyThrLeuGluPr	97
263	CGCCGCCCTCCTGGGTCTCTCCGCTGCCTACCTGTTTCATGCAATGGAGCAGCTGGAGCC	322
97	oAlaAlaGluTrpSerValLeuLeuGlyValHisSerGlnAspGlyProLeuAspGlyAl	117
323	CGCGCGCGAGTGGTCCGTTACTGCTGGCGCTGCATCTCCAGGACGGGCCCTGTGACGGGCG	382
117	aHisThrArgAlaValAlaAlaIleValValProAlaAsnTyrSerGlnValGluLeuG	137
383	GCACACCGCGCAGTGGCGGCCCATCTGCTGCCGCCAACTACAGCAAGTGGAGCTGGG	442
137	yAlaAspLeuAlaLeuLeuArgLeuAlaSerProAlaSerLeuGlyProAlaValTrpPr	157
443	CGCGCGACCTGGCCCTCTGGCGCTGGCCCTACCCGCCAGCCTGGGCCGCCCGCTGTGGCC	502
157	oValCysLeuProArgAlaSerHisSArgPheValHisGlyThrAlaCysTrpAlaThrGl	177
503	TGTCCTGGCTGCCCGCGCCCTCACACGCTTCGTGCACGGCACCCGCTGCTGGGCCACCGG	562
177	yTrpGlyAspValGlnGluAlaAspProLeuProLeuProTrpValLeuGlnValG	197
563	CTGGGGAGAGCTCCAGGAGGACAGATCCTCTGCCTCTCCCTGGGTGTCTACAGGAAGTGA	622
197	uLeuArgLeuLeuGlyGluAlaThrCysGlnCysLeuTyrSerGlnProGlyProPheAs	217
623	GCTAAGGCTGTGGCGCAGGCGCACCTGTCAATGTCTCTACAGCCAGCCCGCTCCCTTCAA	682
217	nLeuThrLeuGlnIleLeuProGlyMetLeuCysAlaGlyTyrProGluGlyArgArgAs	237
683	CCTCACTCTCCAGATATTGCCAGGGATGCTGTGTGCTGGCTACCCAGAGGGCCCGAGGA	742
237	pThrCysGlnGlyAspSerGlyIProProLeuValCysGluGluGlyArgTrpPheGl	257
743	CACCTGCCAGGGCGACCTCTGGGGGGCCCTGGTCTGTGAGGAAGCGCGCCCTGGTTCCA	802
257	nAlaGlyIleThrSerPheGlyPheGlyCysGlyArgArgAsnArgProGlyValPheTh	277
803	GGCAGGAATCAACAGCTTTGGCTTTGGCTGTGGACGGAGAAACCGCCCTGGAGTTTTCAC	862
277	rAlaValAlaThrTyrGluAlaTrpIleArgGluGlnValMetGlySerGluProGlyPr	297
863	TGCTGTGGCTACCTATGAGGCATGATACGGGAGCAGGTGATGGGTTCAGAGCCTGGGCC	922
297	oAlaPheProThrGlnProGlnLysThrGlnSerAsp	309
923	TGCCCTTTCCACCCACGAGCCCGCAAGAACCCAGATCAGAT	959

```

RESULT 7
LOCUS AX342934
DEFINITION Sequence 1 from Patent WO0198467.
ACCESSION AX342934
VERSION AX342934.1
KEYWORDS Homo sapiens (human)
SOURCE Homo sapiens
ORGANISM Enkaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1
AUTHORS Xiao Y. and Morozov, V.
TITLE Regulation of human prostasin-like serine protease
JOURNAL Patent: WO 0198467-A 1 27-DEC-2001;
Bayer Aktiengesellschaft, (DE)
FEATURES
source
1. 537
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/Ab_xref="traxon:9506"

```

ORIGIN

Alignment Scores:	1.28e-44	Length:	537
Pred. No.:	970.00	Matches:	178
Score:	100.00%	Conservative:	0
Percent Similarity:	100.00%	Mismatches:	0
Best Local Similarity:	100.00%	Indels:	0
Query Match:	49.67%	Gaps:	0
DB:	6		
US-10-037-417-46 (3-357) x AX342934 (1-537)			
QY	76	LeulleAlaProSerTrpValLeuSerAlaAlaHisCysPheMetThrAsnGlyThrLeu	95
Db	536	CTCATCGCCCTCTCTGGTCTCTCCGCTGTCTCACTGTTTCATGACGAATGGAGCGT	477
QY	96	GlupProAlaAlaGluTrpSerValLeuLeuGlyValHisSerGlnAspGlyProLeuAsp	115
Db	476	GAGCCCGCGCGAGTGTGGTACTGTCTGGCGTGCATCTCCAGGACGCGGCCCTGGAC	417
QY	116	GlyAlaHisThrArgAlaValAlaAlaIleValProAlaAsnTyrSerGlnValGlu	135
Db	416	GGCGCGCACACCCGCGCAGTGGCGCGCATCTGTGTGGCGGCAACTACAGCCAAAGTGGAG	357
QY	136	LeuGlyAlaAspLeuAlaLeuArgLeuAlaSerProAlaSerLeuGlyProAlaVal	155
Db	356	CTGGCGCGCGACCTGGCCCTGTCTGCCTGTGGCTCATCCCGCCAGCCTGGGCCCCCGCCGTG	297
QY	156	TrpProValCysLeuProArgAlaSerHisArgPheValHisGlyThrAlaCysTrpAla	175
Db	296	TGGCCTGTGTCTGGCCCCCGGCTCTACACGCTTCGTGCACGCAACGCTGTGGGCC	237
QY	176	ThrGlyTrpGlyAspValGlnGluAlaAspProLeuProLeuProTrpValLeuGlnGlu	195
Db	236	ACCGGCTGGGGAGCGTCCAGGAGGAGATCTCTGTGCTCTCCCTGGGTGTACAGAA	177
QY	196	ValGluLeuArgLeuLeuGlyGluAlaThrCysGlnCysLeuTyrSerGlnProGlyPro	215
Db	176	GTGAGAGCTAAGGCTGTCTGGGCGAGGCCACCTGTCAATGTCTCTACAGCCAGCCCCGTCCC	117
QY	216	PheAsnLeuThrLeuGlnIleLeuProGlyMetLeuCysAlaGlyTyrProGluGlyArg	235
Db	116	TTCAACCTCACTCTCCAGATATTGCCAGGATGTCTGTGTGGTATCCAGAGGGCGCG	57
QY	236	ArgAspThrCysGlnGlyAspSerGlyGlyProLeuValCysGluGluGlyGly	253
Db	56	AGGAGACACTGCGCAGGGTGACTCTGGGGGGGGCCCCCTGGTCTGTGAGAAAGCGCGC	3

RESULT 8	BD125219	670 bp	DNA	linear	PAT 18-SEP-2002
LOCUS	BD125219				
DEFINITION	Primer for synthesizing	full-length cDNA			and use thereof.
ACCESSION	BD125219				
VERSION	BD125219.1				GI:23220164
KEYWORDS	JP 2002017375-A/550.				
	homo sapiens (human)				

SOURCE	Homo sapiens
ORGANISM	Homo sapiens
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE	1. (bases 1 to 670)
AUTHORS	Ota.T., Nishikawa,T., Isogai,T., Hayashi,K., Ishii,S., Kawai,Y., Kagamatsu,A., Sugiyama,T., Nagai,K., Kojima,S., Otsuki,T. and Koga,H. for synthesizing full-length cDNA and use thereof
TITLE	Primer for synthesizing full-length cDNA and use thereof
JOURNAL	Patent: JP 2002017375-A 650 22-JAN-2002;
	HELEX RESEARCH INSTITUTE
COMMENT	OS Homo sapiens (human) PN JP 2002017375-A/650 PD 22-JAN-2002 PF 07-JUL-2000 JP 2000253172 PI TOSHIO OTA, TETSUO NISHIKAWA, TAKAO ISOGAI, KOJI HAYASHI, SHIZUKO PI ISHII, PI YURI KAWAI, AI WAKAMATSU, TOMOYASU SUGIYAMA, KEIICHI NAGAI, PI

SHINICHI KOJIMA,
PI TETSUJI OTSUKI, HISASHI KOGA
PC
C12N15/09, C07K14/47, C07K16/18, C12N1/15, C12N1/19, C12N1/21, C12N5/00 CC
PC, C12P21/02, C12Q1/68//C12P21/08, G06F17/30, C12N15/00, C12N5/00 CC
Primer for synthesizing full-length cDNA and use thereof FH Key
Location/Qualifiers
FT source 1..670
FT /organism='Homo sapiens (human)'.
FEATURES
source
Location/Qualifiers
1..670
/organism='Homo sapiens'
/mol_type='genomic DNA'
/db_xref='taxon:9606'
ORIGIN
Alignment Scores: 1.27e-38 Length: 670
Pred. No.: 861.00 Matches: 167
Score: 94.92% Conservative: 1
Percent Similarity: 94.35% Mismatches: 5
Best Local Similarity: 44.09% Indels: 4
Query Match: 6 Gaps: 0
DB:
US-10-037-417-46 (1-357) x BD125219 (1-670)
QY 38 CysGlyArgProGluProSerAlaAargIleVal-GlyGlySerAsnAlaGlnProGlyTh 57
Db 143 TGGGGGGCCCTGAGCCCTCGGCCCGCATCTGGGGGGGGCTCAAAACGGCGAGCGGGCAC 202
QY 57 rTrpProTrrGlnValSerLeuHisHisGlyGlyGlyHisIleCysGlyGlySerLeu 77
Db 203 CTGGCCCTTGGCAAGTGGAGCTGCACCATGGAGGTGGCCACATCTGCGGGGCTCCCTCAT 262
QY 77 eAlaProSerTrpValLeuSerAlaAHisCysPheMetThrAsnGlyThrLeuGluPr 97
Db 263 CGCCCCCTCCCTGGGTCTCTCCGCTGCTCACTGTTTCATGACGATGGAGCTGGAGCC 322
QY 97 oAlaAalaGluTrpSerValLeuGlyValHisSerGlnAspGlyProLeuAspGlyAl 117
Db 323 CGCGGGCGAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 382
QY 117 aHisThrArgAlaValAlaAlaIleValProAlaAsnTrpSerGlnValGluLeuG 137
Db 383 GCACACCGCGAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 442
QY 137 yAlaAspLeuAlaLeuLeuAlaSerProAlaSerLeuGlyProAlaValTrpPr 157
Db 443 CGCGGACCTGGCCCTTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 502
QY 157 oValCysLeuProArgAlaSerHisArgPheValHisGlyThrAlaCysTrpAlaThrGl 177
Db 503 TGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 562
QY 177 yTrpGlyAspVal-GlnGluAlaAspProLeuProLeuProTrpValLeuGln-GluVal 196
Db 563 CTGGGGAGACNTCCANGGAGGAGAAATTCCTGCCCTCTCCCTGGGTGCTTACAGGAAGTG 622
QY 197 GluLeuArgLeuLeuGlyGly-AlaThrCysGlnCysLeuTyr 210
Db 623 GAGCTAAGGNTGCTGGGCGAGGCAACTGGTCAATGTCCTAC 665
RESULT 9
BD126485 670 bp DNA linear PAT 18-SEP-2002
LOCUS
DEFINITION
Primer for synthesizing full-length cDNA and use thereof.
ACCESSION
BD126485
VERSION
BD126485.1 GI:23221430
KEYWORDS
JP 2002017375-A/1916.
SOURCE
Homo sapiens (human)
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata, Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
1 (Bases 1 to 670)
Ota, T., Nishikawa, T., Isogai, T., Hayashi, K., Ishii, S., Kawai, Y.,
Wakamatsu, A., Sugiyama, T., Nagai, K., Kojima, S., Otsuki, T. and
Koga, H. for synthesizing full-length cDNA and use thereof
Patent: JP 2002017375-A 1916 22-JAN-2002;
HELIX RESEARCH INSTITUTE
OS Homo sapiens (human)
PN JP 2002017375-A/1916
PD 22-JAN-2002
PF 07-JUL-2000 JP 2000253172
PI TOSHIO OTA, TETSUO NISHIKAWA, TAKAO ISOGAI, KOJI HAYASHI, SHIZUKO
PI TOSHIKI KAWAI, AI WAKAMATSU, TOMOYASU SUGIYAMA, KEIICHI NAGAI, PI
SHINICHI KOJIMA,
PI TETSUJI OTSUKI, HISASHI KOGA
PC
C12N15/09, C07K14/47, C07K16/18, C12N1/15, C12N1/19, C12N1/21, C12N5/00 CC
PC, C12P21/02, C12Q1/68//C12P21/08, G06F17/30, C12N15/00, C12N5/00 CC
Primer for synthesizing full-length cDNA and use thereof FH Key
Location/Qualifiers
FT source 1..670
FT /organism='Homo sapiens (human)'.
FEATURES
source
Location/Qualifiers
1..670
/organism='Homo sapiens'
/mol_type='genomic DNA'
/db_xref='taxon:9606'
ORIGIN
Alignment Scores: 1.27e-38 Length: 670
Pred. No.: 861.00 Matches: 167
Score: 94.92% Conservative: 1
Percent Similarity: 94.35% Mismatches: 5
Best Local Similarity: 44.09% Indels: 4
Query Match: 6 Gaps: 0
DB:
US-10-037-417-46 (1-357) x BD126485 (1-670)
QY 38 CysGlyArgProGluProSerAlaAargIleVal-GlyGlySerAsnAlaGlnProGlyTh 57
Db 143 TGGGGGGCCCTGAGCCCTCGGCCCGCATCTGGGGGGGGCTCAAAACGGCGAGCGGGCAC 202
QY 57 rTrpProTrrGlnValSerLeuHisHisGlyGlyGlyHisIleCysGlyGlySerLeu 77
Db 203 CTGGCCCTTGGCAAGTGGAGCTGCACCATGGAGGTGGCCACATCTGCGGGGCTCCCTCAT 262
QY 77 eAlaProSerTrpValLeuSerAlaAHisCysPheMetThrAsnGlyThrLeuGluPr 97
Db 263 CGCCCCCTCCCTGGGTCTCTCCGCTGCTCACTGTTTCATGACGATGGAGCTGGAGCC 322
QY 97 oAlaAalaGluTrpSerValLeuGlyValHisSerGlnAspGlyProLeuAspGlyAl 117
Db 323 CGCGGGCGAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 382
QY 117 aHisThrArgAlaValAlaAlaIleValProAlaAsnTrpSerGlnValGluLeuG 137
Db 383 GCACACCGCGAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 442
QY 137 yAlaAspLeuAlaLeuLeuAlaSerProAlaSerLeuGlyProAlaValTrpPr 157
Db 443 CGCGGACCTGGCCCTTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 502
QY 157 oValCysLeuProArgAlaSerHisArgPheValHisGlyThrAlaCysTrpAlaThrGl 177
Db 503 TGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 562
QY 177 yTrpGlyAspVal-GlnGluAlaAspProLeuProLeuProTrpValLeuGln-GluVal 196
Db 563 CTGGGGAGACNTCCANGGAGGAGAAATTCCTGCCCTCTCCCTGGGTGCTTACAGGAAGTG 622
QY 197 GluLeuArgLeuLeuGlyGly-AlaThrCysGlnCysLeuTyr 210
Db 623 GAGCTAAGGNTGCTGGGCGAGGCAACTGGTCAATGTCCTAC 665

QY	197	GluLeuArgLeuLeuGlyGlu-AlaThrCysGlnCysLeuTyr	210
DB	623	GAGCTAAGGNTGCTGGCGAAAGCAACTGTGTCATGTTCTAC	665
 RESULT 10 AX342936/c			
LOCUS	AX342936	456 bp	DNA linear PAT 12-JAN-2002
DEFINITION	Sequence 3 from Patent WO0198467.		
ACCESSION	AX342936		
VERSION	AX342936.1	GI:18152214	
KEYWORDS	Homo sapiens (human)		
SOURCE	Homo sapiens		
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.		
REFERENCE	1	Xiao, Y. and Morozov, V. Regulation of human prostatic-like serine protease Patent: WO 0198467-A 3 27-DEC-2001; Bayer Aktiengesellschaft (DE) Location/Qualifiers	
AUTHORS	1. .456	/organism="Homo sapiens"	
TITLE		/mol_type="unassigned DNA"	
JOURNAL		/db_xref="taxon:9606"	
FEATURES	source		
ORIGIN	Alignment Scores:		
Pred. No.:	2.04e-35	Length:	456
Score:	799.00	Matches:	147
Percent Similarity:	97.37%	Conservative:	1
Best Local Similarity:	96.71%	Mismatches:	4
Query Match:	40.91%	Indels:	0
DB:	6	Gaps:	0
 US-10-037-417-46 (1-357) x AX342936 (1-456)			
QY	97	ProAlaIadlTrpSerValLeuLeuGlyValHisSerGlnAspGlyProLeuAspGly	116
DB	456	CCC GCG CGC AGT GGT CGG TACT GTGG C GNTGC ATCTCC AAGAC GGG CCC CTG GAC GGC	397
QY	117	AlaHisThrArgAlaValAlaAlaIleValValProAlaAsnTyrsSerGlnValGluLeu	136
DB	396	GCGCACACC CGCNGCAGTGCGGCCCATCTG TGTCGCGCCCAA CTACAGCAAGTGGAGCTG	337
QY	137	GlyAlaAspLeuAlaLeuLeuArgLeuAlaSerProAlaSerleucGlyProAlaValTrp	156
DB	336	GCGCGCGACCTCGGCCCTGCTCGCGCTCGCCCTCACCGCTCGTGACGCGCCCGCGGTGG	277
QY	157	ProValCysLeuProArgAlaSerHisArgPheValHisGlyThrAlaCysTrpAlaThr	176
DB	276	CCTGTCTGCTTGCCCGCGCCCTCACCGCTCGTGACGCGCCCGCGGTGG	217
QY	177	GlyTrpGlyAspValGlnGluAlaAspProLeuProLeuProTrpValLeuGlnGluVal	196
DB	216	GGCTGGGAGAGCTCCAGAGGACAGATCCTCTGCTCTCCCTGGGTGCTACAGGAAGTG	157
QY	197	GluLeuArgLeuLeuGlyGluAlaThrCysGlnCysLeuTyrSerGlnProGlyProPhe	216
DB	156	GAGCTAAGGCTGCTGGCGAGGCCACCTGTCAATGCTCTCTACAGCCAGCCGCGTCCCTTC	97
QY	217	AsnLeuThrLeuGlnIleLeuProGlyMetLeuCysAlaGlyTyrProGluGlyArgArg	236
DB	96	AACCTCACTCTCCAGATATTGCAGGGATGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT	37
QY	237	AspThrCysGlnGlyAspSerGlyGlyProLeuVal	248
DB	36	GACACCTGCCAGGAGTGAATCTGGGGGGCCCCCTGGTC	1
 RESULT 11 AC009088/c			
LOCUS	AC009088	127769 bp	DNA linear PRI 29-MAR-2003

Qy 78 AlaProSerTrpValLeuSerAlaAlaHisCysPheMetThr----- 91
Db 32028 GCCCCTCTCGGGTCTCTCGCTGCTCACTGTTTCATGACGTAGTATCGCCTGTC 31969
Qy 91 ----- 91
Db 31968 CTGGCCCTCTCGGCCAAAGAGTGGCCCCCGCCCGCAGTTGTTCCCTCTCTCGCGCC 31909
Qy 91 ----- 91
Db 31908 CACTTCCTGCTCAAGGATAAATTGGAGCAAGCTGTCCCAAGCGAGTGTCTATTGCA 31849
Qy 91 ----- 91
Db 31848 GGCATCATGTTCCAGTGGGCAAGGTGTCGGAGCTGGTCCACCACTCCACTCAATAT 31789
Qy 91 ----- 91
Db 31788 CCGGGTCTGAATCCCATCTTTCCATGACCCCGCCACACAGCTCCAAATCGGATTTCG 31729
Qy 91 ----- 91
Db 31728 TCTTCCCTTTGAGGAGTATCAGACCAAAATCTGGATCGGATAGCACCTGCCCTCTGGC 31669
Qy 91 ----- 91
Db 31668 TCTGAGCCCTGCCCCCAACAGGGTCTGCCCTCTCATAGGTCAACCTAACCTCTAC 31609
Qy 92 -----AsnGlyThrLeuGluProAlaAlaAlaGluTrpSerValLeu 105
Db 31608 TGCCCGCTCGAGCGAGGATGGAGCTGGAGCCCGCGCGGCGACACCGCGAGTGGCGCAT 31549
Qy 105 uGlyValHisSerGlnAspGlyProLeuAspGlyAlaHisThrArgAlaValAlaAla 125
Db 31548 GGGGTGCACTCCACAGACGGGCGCTGGAGCGGCGCGACACCGCGAGTGGCGCAT 31489
Qy 125 eValValProAlaAsnTyrSerGlnValGluLeuGlyAlaAspLeuAlaLeuArgLe 145
Db 31488 CGTGGTCCGCGCAACATACAGCAAGTGGAGCTGGGCGCGCACTGGCCCTGTGGCGCT 31429
Qy 145 uAlaSerProAlaSerLeuGlyProAlaValTrpProValCysLeuProArgAlaSerHi 165
Db 31428 GGCCTCACCCCGAGCTGGCGCGCGCTGTGCTGTGCTGCTGCTGCTGCTGCTGCTGCT 31369
Qy 165 sArgPheValHisGlyThrAlaCysTrpAlaThrGlyTrpGlyAspValGlnGluAlaAs 185
Db 31368 CGCTTCTGTGACGACCGGCTGTGCGGCCACCGGCTGGGAGACGFTCCAGGAGGAGG 31309
Qy 185 pProLeuProLeuProTrp-----ValLeuGlnGluVal----- 196
Db 31308 T-----GAGTGGGACGGACAGAGGGAGGGCTCGGGGAGGTGGGGCTTCG 31261
Qy 197 -GluLeuArgLeuLeuGly 202
Db 31260 AGAGTGAGGTTCCTAGGG 31242

RESULT 12
AX098193
LOCUS AX098193
DEFINITION Sequence 105 from Patent WO0118542.
ACCESSION AX098193
VERSION AX098193.1 GI:13515276
KEYWORDS Homo sapiens (human)
SOURCE
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1
Lee, J., Thompson, P. and Lillie, J.
REFERENCE
AUTHORS Identification, assessment, prevention, and therapy of ovarian
TITLE cancer
JOURNAL Patent: WO 0118542-A 105 15-MAR-2001;
Millennium Predictive Medicine, Inc. (US)

FEATURES
source
Location/Qualifiers
1..1796
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"
ORIGIN
Alignment Scores: 1796
Pred. No.: 2,4e-29 Length: 1796
Score: 697.00 Matches: 160
Percent Similarity: 55.29% Conservative: 49
Best local Similarity: 42.33% Mismatches: 133
Query Match: 35.69% Indels: 36
DB: 6 Gaps: 11
US-10-037-417-46 (1-357) x AX098193 (1-1796)
Qy 1 MetAlaGlnValGlyValLeuGlyProGlyGlnLeuGlyAlaValAlaAsnSerAspSer 20
Db 152 ATGGCCCAAGAGGGGGTCTCTGGGGCTGGGAGCTGGGGCTGTGGCC----- 199
Qy 21 TyrSerLeuTyr---GlyLeuValProSerGlyPro-----AlaArgGlyPro 35
Db 200 ATTCTGCTCTATCTTGGATTACTCGGTGGGGACAGAGCGGAAAGGGGCGAAGCTCCC 259
Qy 36 ProTyrCysGlyArgProGluProSerAlaArgLeValGlyGlySerAsnAlaGlnPro 55
Db 260 -----TGCGGT---GTGGCCCCCAAGCACCATCATGAGAGGGCTCCATGTGTGGTGTCT 310
Qy 56 GlyThrTrpProTrpGlnValSerLeuHisGlyGlyGlyHisLeCysGlyGlySer 75
Db 311 GGTACGTGGCCCTGGCAGGTGAGTACCTATGAGAGGGCTCCATGTGTGTGGTGTCT 370
Qy 76 LeuLeAlaProSerTrpValLeuSerAlaAlaHisCysPheMetThrAsnGlyThrLeu 95
Db 371 CTGCTGCTGAGCAGTGGGTGCTGTCTGAGTCTCTCTCTCTCTCTCTCTCTCTCTCT 430
Qy 96 GluProAlaAlaGluTrpSerValLeuLeuGlyValHisSerGlnAspGlyProLeuAsp 115
Db 431 GAA-----GCCATGAGGTCAAGCTGGGGGCCACCCAGCTAGACTCTCTCTCTCTCTCT 481
Qy 116 GlyAlaHisThrArgAlaValAlaAlaValValProAlaAlaSerGlnValGlu 135
Db 482 GAGCCAGGTGAGCCCTGAGGACATCATCCCCACCCAGCTACCTCTCCAGAGGGC 541
Qy 136 LeuGlyAlaAspLeuAlaLeuLeuArgLeuAlaSerProAlaSerLeuGlyProAlaVal 155
Db 542 TCCAGGGCGACATTGCACTCTCTCAACTCAGCAGACCCCATCCTCTCTCTCTCTCTCTCT 601
Qy 156 TrpProValCysLeuProArgAlaSerHisArgPheValHisGlyThrAlaCysTrpAla 175
Db 602 CGGCCCATCT 661
Qy 176 ThrGlyTrpGlyAspValGlnGluAlaAspProLeuProLeuProTrpValLeuGlnGlu 195
Db 662 ACTGGTGGGGTCTATGTGGCCCTCTGAGTGGCTCTCTCTCTCTCTCTCTCTCTCTCTCT 721
Qy 196 ValGluLeuArgLeuLeuGlyGluAlaThrCysGlnCysLeuTyrSerGlnProGlyPro 215
Db 722 CTCGAGGTGCTCTGATCAGTCTGAGACGTGTAACTGCTGTATGCTGTGTGTGTGTGTGT 781
Qy 216 PheAsnLeuThrLeuGlnLeuProGlyMetLeuCysAlaGlyTyrProGluGlyArg 235
Db 782 CCTGAGGAGCGCATTTGTCTCAAGAGGACATGGTGTGTGTGTGTGTGTGTGTGTGTGT 841
Qy 236 ArgAspThrCysGlnGlyAspSerGlyGlyProLeuValCysGluGluGlyClyArgTrp 255
Db 842 AAGGACGCTGCGCAGGTGACTCTGGGGGCCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 901
Qy 256 PheGlnAlaGlyLeuThrSerPheGlyPheGlyCysGlyArgArgAsnArgProGlyVal 275
Db 902 TACCTGACGGGCATTGTGAGCTGGGAGATGCTGTGGGGGCCCGCCACAGGCTGTGTGTGT 961

```

Web site: http://www.nisac.nih.gov/
Contact: nisc_mg@nih.gov
Akhter, N., Ayele, K., Beckstrom-Sternberg, S.M., Benjamin, B.,
Blakesley, R.W., Bouffard, G.G., Breen, K., Brinkley, C., Brooks, S.,
Dietrich, N.L., Granite, S., Guan, X., Gupta, J., Hagnighi, P.,
Hansen, N., Ho, S.-L., Karlins, E., Kwong, P., Lalic, P., Legaspi, R.,
Maduro, Q.L., Vasiello, C., Maskeri, B., Mastrian, S.D., McCloskey, J.C.,
McDowell, J., Pearson, R., Stantripop, S., Thomas, P.J., Touchman, J.W.,
Turgeon, C., Vogt, J.L., Walker, M.A., Wetherby, K.D., Wiggins, L.,
Young, A., Zhang, L.-H. and Green, E.D.

Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
Series: IRAL Plate: 4 Row: j Column: 3
This clone was selected for full length sequencing because it
passed the following selection criteria: matched mRNA gi: 21536453.
Location/Qualifiers
1. 1800
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="MGC:2133 IMAGE:3138532"
/tissue_type="Placenta, choriocarcinoma"
/clone_lib="NIH_MGC_21"
/lab_host="DH10B-R"
/note="Vector: pOTB7"
1. 1800
/gene="PRSS8"
/note="synonym: PROSTASIN"
/db_xref="locusid:5652"
/db_xref="MIM:600823"
176. 1207
/codon_start=1
/product="prostatic preproprotein"
/protein_id="AAH01462.1"
/db_xref="GI:12655207"
/db_xref="locusid:5652"
/translation="MAQVGLGPGQLGAVALLVGLLSGTGAEGAEAPCGVAPCAR
ITGSSAVAGPMQVSVITVEGVHVGGSIVSGLVSAHCFPPSEPHKEAYEVKLGA
HQLDSYSDAKVSTLKDIIIPHSYLOQSGQDIALQLSKPTFTSYRIPICLPANA
SPNGLCHTQVGHVPSVSLTPKPLQQLVPLASRETCLNYINDAKPEPHFVAS
EDMGCAVYGGKACQDGSGLSPCPVEGLVLTGIVSGWDGAGARNPGVTVLASS
YASVLSQKVTLEQPRVVPQTQESPDNSLQSHLAFSSAPAQGLLRPLFLPLGLALG
LUSPWLSEH"
misc_feature
305. 1018
/note="Tryp Spc; Region: Trypsin-like serine protease"
/db_xref="CDD:smart00020"

ORIGIN

Alignment Scores:
2,41e-29 Length: 1800
697.00 Matches: 160
Score: 55.29% Conservative: 49
Percent Similarity: 42.33% Mismatches: 133
Best Local Similarity: 35.69% Indels: 36
Query Match: 9 Gaps: 11
DB:

US-10-037-417-46 (1-357) x BC001462 (1-1800)
QY 1 MetAlaGlnLysGlyValLeuGlyProGlyGlnLeuGlyAlaValAlaAsnSerAspSer 20
Db 176 ATGCCCCAGAGGGGGTCTGGGGCCCTGGGCAGCTGGGGGCTGTGCC----- 223
QY 21 TyrSerLeuTyr---GlyLeuValProSerGlyPro-----AlaArgGlyPro 35
Db 224 ATTCTGCTCTATCTGTGATTACTCCGCTGGGGACAGGCGGAGGGGCAGAGCTCCC 283
QY 36 ProTyrCysGlyArgProGluProSerAlaArgIleValGlyGlySerAsnAlaGlnPro 55
Db 284 -----TCCGGT---GTGGCCCCCAAGCAGCGCATCACAGGTGGCAGCAGTCACTGCC 334
QY 56 GlyThrTrpProTrpGlnValSerLeuHisGlyGlyGlyHisIleCysGlyGlySer 75

```



```

Db      919  AAGACGCGCTGCCAGGCTGACTCTGGGGGCCCACTCTCCCTGCGCTGTGGAGGCTCTCTGG 978
QY      256  PheGlnAlaGlyIleThrSerPheGlyPheGlyCysGlyArgArgAsnArgProGlyVal 275
Db      979  TACCTGACGGGCATTTGAGCTGGGGAGATGCCGTGTGGGGCCCGCAACAGGCTGGTGTG 1038
QY      276  PheThrAlaValAlaThrTyrGluAlaTyrPheArgGluGlnValMetGlySerGluPro 295
Db      1039  TACACTCTGGCTCCAGCTATGCTCTGATCCATGATCCAAAGCAAGTG-----ACAGAACTC 1092
QY      296  GlyProAlaPheProThrGlnProGlnLysThrGlnSerAsp-----CysLeuHis 312
Db      1093  CAGCGCTCGTGTGGTGGCCCAACACAGAGTCCAGGCCCGACACCACTCTGTGGCAGC 1152
QY      313  GlnThrAlaPheLeuAspSer---AlaArgIleLeuLeuArgProLeuSerHisIleSer 331
Db      1153  CACCTGGCTTCAGCTCTGCCCGACGCCAGGGCTTGCTGAGGCCCATCTCTTTCTGCT 1212
QY      332  ValGlyValSerThrGlyThrLysSerLeuValLeuProTyrLeuSer----- 347
Db      1213  CTGGCGCTGGCTCTGGC-----CTGCTCTCCCATGCTCAGCGAGCACTGAGCT 1263
QY      348  -----ProHisSerLeuLeuGlyLeuTyrGlyPhe 357
Db      1264  GGCCCTACTTCCAGGATGATGCATCACACTCAAGGACAGGAGCGCTGCTCTTC 1317

RESULT 15
AX336076 1834 bp DNA linear PAT 09-JAN-2002
LOCUS AX336076 Sequence 6585 from Patent WO0194629.
DEFINITION AX336076
ACCESSION AX336076
VERSION AX336076.1 GI:18126795
KEYWORDS Homo sapiens (human)
SOURCE Homo sapiens
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1
AUTHORS Young, P. E., Augustus, M., Carter, K. C., Ebner, R., Endress, G.,
Horrigan, S., Soppet, D. R. and Weaver, Z.
TITLE Cancer gene determination and therapeutic screening using signature
gene sets
JOURNAL Patent: WO 0194629-A 6585 13-DEC-2001;
Avalon Pharmaceuticals (US)
FEATURES
source
1..1834
Location/Qualifiers
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"
ORIGIN
Alignment Scores:
Pred. No.: 2,45e-29 Length: 1834
Score: 697.00 Matches: 160
Percent Similarity: 55.2% Conservative: 49
Best local Similarity: 42.33% Mismatches: 133
Query Match: 35.69% Indels: 36
DB: 6 Gaps: 11

US-10-037-417-46 (1-357) x AX336076 (1-1834)
QY 1 MetAlaGlnLysGlyValLeuGlyProGlyGlnLeuGlyAlaValAlaAsnSerAspSer 20
Db 229 ATGGCCCAAGAGGGGTCTCTGGGGCTGGGAGCTGGGGGCTGTGGCC----- 276
QY 21 TyrSerLeuTyr---GlyLeuValProSerGlyPro-----AlaArgGlyPro 35
Db 277 ATTCTGCTCTATCTTGGATTACTCCGGTCTGGGACAGAGGCGGAGGGGCGAGAGCTCC 336
QY 36 ProTyrCysGlyArgProGluProSerAlaArgIleValGlySerAsnAlaGlnPro 55
Db 337 -----TGCGGT---GTGGCCCCCAAGCAGCGATCACAGGTGGCAGCTGAGTCGCGC 387

```

Search completed: February 25, 2004, 17:20:01
Job time : 3826 secs

```

QY 56 GlyThrTrpProTyrGlnValSerLeuHisHisGlyGlyGlyHisIleCysGlyGlySer 75
Db 388 GGTCAAGTGGCCCTGGCAGCTCAGCTACCTATGAAGGCGTCCATGTGTGTGTGTGTGTGT 447
QY 76 LeuIleAlaProSerTrpValLeuSerAlaAlaHisCysPheMetThrAsnGlyThrLeu 95
Db 448 CTGCTGTGTGAGCAGTGGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 507
QY 96 GluProAlaAlaGlnTrpSerValLeuLeuGlyValHisSerGlnAspGlyProLeuAsp 115
Db 508 GAA-----GCCTATGAGGTCAAGCTGGGGGCCCAACAGCTCAGCTCTCTCTCTCTCT 558
QY 116 GlyAlaHisThrArgAlaValAlaAlaIleValProAlaAsnTyrSerGlnValGlu 135
Db 559 GAGGCCAAGCTCAGCACCTCTGAAGGACATCATCCCCACCCAGCTACCTCTCAGGAGGCG 618
QY 136 LeuGlyAlaAspLeuAlaLeuLeuArgLeuLeuAspProAlaSerLeuGlyProAlaVal 155
Db 619 TCCAGGGCGCAGCATGTGACTCTCCAACTCAGCAGACCCATCACCTTCTCCCGCTACATC 678
QY 156 TrpProValCysLeuProArgAlaSerHisArgPheValHisGlyThrAlaCysTrpAla 175
Db 679 CGGCCCATCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 738
QY 176 ThrGlyTrpGlyAspValGlnGluAlaAspProLeuProLeuProTyrValLeuGlnGlu 195
Db 739 ACTGGCTGGGTCTGTGGCCCTCTGAGTGTCTCTGAGCTCTCTGAGCCCAAGCCACTGCAG 798
QY 196 ValGlnLeuArgLeuLeuGlyGluAlaThrCysGlnCysLeuTyrSerGlnProGlyPro 215
Db 799 CTCGAGGTGCTCTGATCAGTCGTGTGAGACGTGTAACTGCCTGTACACATCGACGCCAAG 858
QY 216 PheAsnLeuThrLeuGlnIleLeuProGlyMetLeuCysAlaGlyTyrProGluGlyArg 235
Db 859 CCTGAGGAGCGCCTTTGTCCAGAGGACATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 918
QY 236 ArgAspThrCysGlnGlyAspSerGlyGlyProLeuValCysGlnGluGlyArgTrp 255
Db 919 AAGGAGCGCTGCCAGGTGACTCTGGGGGCCCACTCTCTGCTGTGTGTGTGTGTGTGTGTGT 978
QY 256 PheGlnAlaGlyIleThrSerPheGlyPheGlyCysGlyArgArgAsnArgProGlyVal 275
Db 979 TACCTGACGGGCATTTGAGCTGGGGAGATGCCGTGTGGGGCCCGCAACAGGCTGGTGTG 1038
QY 276 PheThrAlaValAlaThrTyrGluAlaTyrPheArgGluGlnValMetGlySerGluPro 295
Db 1039 TACACTCTGGCTCCAGCTATGCTCTGATCCAAAGCAAGTG-----ACAGAACTC 1092
QY 296 GlyProAlaPheProThrGlnProGlnLysThrGlnSerAsp-----CysLeuHis 312
Db 1093 CAGCGCTCGTGTGGTGGCCCAACACAGGAGTCCAGGCCCGACACCACTCTCTGTGGCAGC 1152
QY 313 GlnThrAlaPheLeuAspSer---AlaArgIleLeuLeuArgProLeuSerHisIleSer 331
Db 1153 CACCTGGCTTCAGCTCTGCCCGACGCCAGGGCTTGCTGAGGCCCATCTCTTTCTGCTCT 1212
QY 332 ValGlyValSerThrGlyThrLysSerLeuValLeuProTyrLeuSer----- 347
Db 1213 CTGGCGCTGGCTCTGGC-----CTGCTCTCCCATGCTCAGCGAGCACTGAGCT 1263
QY 348 -----ProHisSerLeuLeuGlyLeuTyrGlyPhe 357
Db 1264 GGCCCTACTTCCAGGATGATGCATCACACTCAAGGACAGGAGCGCTGCTCTTC 1317

```


Result No.	Score	Query		ID	Description
		Match	Length		
1	1953	100.0	1102	ABQ93301	Human pro
2	1953	100.0	1102	ABQ93302	Human pro
3	1505	77.1	2457	ABK31795	DNA encod
4	1505	77.1	2681	ABQ75956	Human PM
5	1492	76.4	2810	AAK94500	Human ful
C	970	49.7	537	ABK13565	Human pro
	861	44.1	670	AAK92190	Human CDN
	861	44.1	670	AAK93456	Human CDN

FT FT /*tag= b
FT FT /partial
FT FT /product= "NOV14a"
XX XX /note= "No stop codon given in the specification"

PN WO200253742-A2.

PN 11-JUL-2002.

XX 07-JAN-2002; 2002WO-US000375.

XX 05-JAN-2001; 2001US-0260018P.

XX 08-JAN-2001; 2001US-0260360P.

XX 28-FEB-2001; 2001US-0272411P.

XX 02-MAR-2001; 2001US-0272817P.

XX 05-JUL-2001; 2001US-0303231P.

XX 12-JUL-2001; 2001US-0305060P.

XX 10-SEP-2001; 2001US-0318405P.

XX 12-SEP-2001; 2001US-0318700P.

XX 04-JAN-2002; 2002US-00037417.

XX (CURA-) CURAGEN CORP.

XX Kekuda R, Alsobrook JP, Tchernev VT, Liu X, Spytek KA;

XX Patturajan M, Grosse WM, Lepley DM, Burgess CE, Vernet CAM, Li L;

XX Gorman L, Edinger S, Sciore P, Ellerman K, Malyankar U;

XX Rothenberg M, Stone D, Boidog F, Guo X, Shenoy S, Anderson D;

XX Padigaru M, Taupier RJ, Miller CE, Eisen A;

XX WPI; 2002-583619/62.

XX P-PSDB; ABH09523.

XX Novel polypeptides and nucleic acids homologous to transmembrane

XX receptor, thymosin, neuromodulin-like family of proteins for diagnosing,

XX treating cancer, atherosclerosis, neurological, skin and autoimmune

XX disorders.

XX Claim 9a; Page 142; 323pp; English.

XX The invention relates to 24 novel human proteins designated NOV1-NOV14

XX (AB09501-AB09524), collectively referred to as NOVX proteins, and

XX nucleic acids encoding them (AB093879-AB093902). NOVX proteins and

XX nucleotides are useful in the treatment, diagnosis or prevention of NOVX-

XX associated disorders or in the manufacture of a medicament for treating

XX such disorders, with specific applications described for each of the 24

XX NOVX proteins, based on their homology to known proteins. Various

XX disorders are associated with NOVX proteins including neurological

XX disorders (e.g., Alzheimer's, Huntington's and Parkinson's diseases),

XX pain, behavioural disorders, addiction, tuberculous sclerosis, cancers

XX (e.g., colorectal cancer, leukaemia and osteosarcoma), immune disorders

XX various forms of arthritis, diabetes, thyroiditis, cardiovascular disease

XX (e.g., hypertension), reproductive disorders, endometriosis,

SQ Sequence 1102 BP; 162 A; 371 C; 353 G; 216 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.: 1.39e-99 Length: 1102

Score: 1953.00 Matches: 357

Percent Similarity: 100.00% Conservative: 0

Best Local Similarity: 100.00% Mismatches: 0

Query Match: 100.00% Indels: 0

DB: 6 Gaps: 0

US-10-037-417-46 (1-357) x.ABQ93901 (1-1102)

QY 1 MetAlaGlnIysGlyValLeuGlyProGlyGlnLeuGlyAlaValAlaAsnSerAspSer 20

Db 19 ATGGCCCAAGAGGGGTCCTGGGGCCCTGGGCAGCTGGGGCTGTGGCAATTCATCTCA 78

QY 21 TysSerLeuTyrGlyLeuValProSerGlyProAlaArgGlyProProTyrCysGlyArg 40

Db 79 TACTCACTTTACGGGTGGTGGTCCGTCAGCCGCTAGGGGGCCCCCGGTATCTGGGGCGC 138

QY 41 ProGluProSerAlaArgIleValGlyGlySerAsnAlaGlnProGlyThrTrpProTrp 60

Db 139 CCTGAGCCCTCGGCCCGCATCGTGGGGGGCTCAAAACGCGCAGCGGCGACCTGGCCCTGG 198

QY 61 GlnValSerLeuHisHisGlyGlyGlyHisIleCysGlyGlySerLeuIleAlaProSer 80

Db 199 CAAGTGAGCCCTGCACCATGGAGGGTGGCCACATCTGGGGGGCTCCCTCATCGCCCTCC 258

QY 81 TrpValLeuSerAlaAlaHisCysPheMetThrAsnGlyThrLeuGluProAlaAlaGlu 100

Db 259 TGGGTCTCTCGCTGCTCACTGTTTTCATGCAATGGGACGTTGGAGCCCGCGCCGAG 318

QY 101 TrpSerValLeuLeuGlyValHisSerGlnAspGlyProLeuAspGlyAlaHisThrArg 120

Db 319 TGGTGGTACTCTGGGGCTGCACCTCCACGAGACGGGGCCCTGGACGGCGCGCACCCGC 378

QY 121 AlaValAlaAlaIleValValProAlaAsnTyrSerGlnValGluLeuGlyAlaAspLeu 140

Db 379 GCAGTGGCCCGCATCGTGGTGGCGGCCACTACAGCAAGTGGAGCTGGGGCGCGACCTG 438

QY 141 AlaLeuLeuArgLeuAlaSerProAlaSerLeuGlyProAlaValTrpProValCysLeu 160

Db 439 GCCTGTCTGGCTGGCTGCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTG 498

QY 161 ProArgAlaSerHisArgPheValHisGlyThrAlaCysTrpAlaThrGlyTrpGlyAsp 180

Db 499 CCGCGGGCCCTCACACCGCTTCGTGACGGGACCCGCTGTGGCCACCGCTGGGGAGAC 558

QY 181 ValGlnGluAlaAspProLeuProLeuProTrpValLeuGlnGluValGluLeuArgLeu 200

Db 559 GTCCAGGAGGACGATCTCTCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTG 618

QY 201 LeuGlyGluAlaThrCysGlnCysLeuTyrSerGlnProGlyProPheAsnLeuThrLeu 220

Db 619 CTGGCGAGGCGCACCTGTCAATGTCTCTACAGCCAGCCCGCTCCCTTCAACCTCACTTC 678

QY 221 GlnIleLeuProGlyMetLeuCysAlaGlyTyrProGluGlyArgArgAspThrCysGln 240

Db 679 CAGATATTGCCAGGATGCTGTGCTGGCTACCCAGAGGGGCGCGAGGACACCTGCCAG 738

QY 241 GlyAspSerGlyGlyProLeuValCysGluGluGlyArgTrpPheGlnAlaGlyIle 260

Db 739 GGTGACTCTGGGGGGCCCTCTGTGTGAGGAAGCGCCGCTGGTTCAGGCGAGGAATC 798

QY 261 ThrSerPheGlyPheGlyCysGlyArgArgAsnArgProGlyValPheThrAlaValAla 280

Db 799 ACCAGCTTTGGGTTTGGCTGTGGACGGAGAACCGCCCTGGAGTTTTCACCTGTGCTGC 858

QY 281 ThrTyrGluAlaTrpIleArgGluGlnValMetGlySerGluProGlyProAlaPhePro 300

Db 859 ACCTATGAGGCGATGATACGGGAGCGAGGTGATGGGTTCAGAGCCTGGGCTGCCCTTCCC 918

QY 301 ThrGlnProGlnLysThrGlnSerAspCysLeuHisGlnThrAlaPheLeuAspSerAla 320

XX

Db 919 ACCGAGCCCAAGACACCTAGATTGTTATCATCAACGCGATCTCTGGATCTGCC 978
Qy 321 ArgileLeuLeuAigProLeuSerHisileSerValGlyValSerThrGlyThrIlySer 340
Db 979 AGAATCCCTTTGAGGCCCTTGTCCATATATACAGTAGAGTCTCAACTGGACCAAAAGC 1038
Qy 341 LeuValLeuProTrpLeuSerProHisSerLeuLeuGlyLeuTrpGlyPhe 357
Db 1039 CTTGTCCTCCCTGGCTCTCTCCACACTCTCTCTCTGGGCGCTCTGGGGGTC 1089
RESULT 2
ID ABQ93902 standard; DNA; 1102 BP.
XX ABQ93902;
AC ABQ93902;
XX 01-NOV-2002 (first entry)
DT Human prostatic precursor-like NOV14b DNA, SEQ ID NO:45.
DE Human; NOVX; neurological disorder; Alzheimer's disease;
XX Huntington's disease; parkinson's disease; pain; behavioural disorder;
KW addition; tuberos scleros; cancer; immune disorder; allergy;
KW autoimmune disease; myasthenia gravis; asthma; arthritis; diabetes;
KW thyroiditis; cardiovascular disease; hypertension; reproductive disorder;
KW endometriosis; cirrhosis; psoriasis; scleroderma; alopecia; ulcer;
KW pancreatitis; incontinence; glomerular endotheliosis; bacterial infection;
KW polycystic kidney disease; endocrine disorder; obesity; cardiomyopathy;
KW atherosclerosis; cell signal processing-related disorder;
KW metabolic pathway regulation disorder; cytostatic; neuroprotective;
KW antiinflammatory; immunosuppressive; analgesic; antiatherosclerotic;
KW dermatological; antibacterial; antarthritic; hepatotropic; neurogenesis;
KW differentiation; proliferation; motility; haematopoiesis; wound healing;
KW angiogenesis; forensic biology; transgenic animal; drug screening;
KW gene therapy; NOV14b; prostatic precursor-like; gene; ds.
XX Homo sapiens.
XX Key Location/Qualifiers
PH 1. .18
FT 5'UTR /tag= a
FT CDS 19. .1092
FT /*tag= b
FT /*product= "NOV14b"
FT 3'UTR 1093. .1102
FT /*tag= c
XX WO200253742-A2.
XX 11-JUL-2002.
XX 07-JAN-2002; 2002WO-US000375.
XX 05-JAN-2001; 2001US-0260018P.
XX 08-JAN-2001; 2001US-0260360P.
XX 28-FEB-2001; 2001US-0272411P.
XX 02-MAR-2001; 2001US-0272817P.
XX 05-JUL-2001; 2001US-0303231P.
XX 12-JUL-2001; 2001US-0305060P.
XX 10-SEP-2001; 2001US-0318405P.
XX 12-SEP-2001; 2001US-0318700P.
XX 04-JAN-2002; 2002US-00037417.
XX (CURA-) CURAGEN CORP.
XX Kekuda R, Alsobrook JP, Tchernev VT, Liu X, Spytek KA;
PI Patturajan M, Grosse WM, Lepley DM, Burgess CE, Vernet CAM, Li L;
PI Gorman L, Edinger S, Sciorio P, Ellerman K, Malvankar U;
PI Rothenberg M, Stone D, Boldog F, Guo X, Shenoy S, Anderson D;
PI Padigaru M, Taupier RJ, Miller CE, Eisen A;
XX WPI; 2002-583619/62.

DR P-PSDB; ABB09524.
XX Novel polypeptides and nucleic acids homologous to transmembrane
PT receptor, thymosin, neuromodulin-like family of proteins for diagnosing,
PT treating cancer, atherosclerosis, neurological, skin and autoimmune
PT disorders.
XX Claim 9a; Page 143; 323pp; English.
XX The invention relates to 24 novel human proteins designated NOV1-NOV14
CC (ABB09501-ABB09524), collectively referred to as NOVX proteins, and
CC nucleic acids encoding them (ABQ93879-ABQ93902). NOVX proteins, and
CC nucleotides are useful in the treatment, diagnosis or prevention of NOVX-
CC associated disorders or in the manufacture of a medicament for treating
CC such disorders, with specific applications described for each of the 24
CC NOVX proteins, based on their homology to known proteins. Various
CC disorders are associated with NOVX proteins including neurological
CC disorders (e.g., Alzheimer's, Huntington's and Parkinson's diseases),
CC pain, behavioural disorders, addiction, tuberos scleros, cancers
CC (e.g., colorectal cancer, leukaemia and osteosarcoma), immune disorders
CC (e.g., allergies and autoimmune diseases), myasthenia gravis, asthma,
CC various forms of arthritis, diabetes, thyroiditis, cardiovascular disease
CC (e.g., hypertension), reproductive disorders, endometriosis,
CC incontinence, psoriasis, scleroderma, alopecia, ulcers, pancreatitis,
CC cirrhosis, glomerular endotheliosis, polycystic kidney disease, endocrine
CC disorders, obesity, bacterial infections and particularly cardiomyopathy,
CC atherosclerosis, cell signal processing-related disorders and disorders
CC of metabolic pathway regulation. NOVX nucleic acids and polypeptides may
CC be used to identify cellular receptors or downstream effectors which
CC binds to a NOVX protein, and are also useful as targets for the
CC identification of small molecules that modulate or inhibit processes such
CC as neurogenesis, cell differentiation, cell motility, cellular
CC proliferation, haematopoiesis, wound healing and angiogenesis. NOVX
CC nucleic acid sequences can be used to identify a cell or tissue type and
CC are useful as a source of primers or probes for forensic biology and for
CC identifying and cloning NOVX homologues in other cell types. Cells
CC comprising NOVX nucleic acids are useful for producing non-human
CC transgenic animals which are useful for studying the function and
CC activity of NOVX proteins and for identifying and evaluating modulators
CC of NOVX activity. The present sequence represents DNA encoding the
CC prostatic precursor-like protein NOV14b
XX Sequence 1102 BP; 162 A; 371 C; 353 G; 216 T; 0 U; 0 Other;
SQ Alignment Scores:
Pred. No.: 1.39e-99 Length: 1102
Score: 1953.00 Matches: 357
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: Gaps: 0
US-10-037-417-46 (1-357) x ABQ93902 (1-1102)
Qy 1 MetAlaGlnIlyGlyValLeuGlyProGlyGlnLeuGlyAlaValAlaAsnSerAspSer 20
Db 19 ATGCCCCAGAGGGGGTCTTGGGGCTGGGCGAGCTGGGGGCTGTGGCCAAATCTGACTCA 78
Qy 21 TyrSerLeuTyrGlyLeuValProSerGlyProAlaArgGlyProProTyrCysGlyArg 40
Db 79 TACTCATTATACGGTGTGGTCCGACCCGCTAGGGGCCCCCGTACTGCGGGGCGC 138
Qy 41 ProGluProSerAlaArgIleValGlyGlySerAsnAlaGlnProGlyThrTrpProTrp 60
Db 139 CCGAGCCCTCGGGCCGCATCGTGGGGGGGCTCAACAGCGCGAGCCGACCTGGCCTTGG 198
Qy 61 GlnValSerLeuHisGlyGlyGlyHisIleCysGlyGlySerLeuIleAlaProSer 80
Db 199 CAAGTGAGCCTGCACATGAGAGGTGGCCACATCTCGGGGGGCTCCCTCATCGCCCCCTCC 258
Qy 81 TrpValLeuSerAlaAlaHisCysPheMetThrAsnGlyThrLeuGluProAlaAlaGlu 100
Db 259 TGGGTCTCTCTCGCTGCTCACTGTTTCATGACGAATGGGACGTTGGAGCCCGCGCGAG 318

QY 101 TrpSerValLeuLeuGlyValHisSerGlnAspGlyProLeuAspGlyAlaHisThrArg 120
Db 319 TGGTGGTACTGCTGGGCGTGCATCCAGGACGGGCCCTGGACGGCGGCACCCGC 378
QY 121 AlaValAlaAlaValValProAlaAsnTyrSerGlnValGluLeuGlyAlaAspLeu 140
Db 379 GCAGTGGCGGCACATCGTGGTGGCGGCACTACAGCCAAAGTGGAGCTGGGCGCGACCTG 438
QY 141 AlaLeuLeuArgLeuAlaSerProAlaSerLeuGlyProAlaValTyrProValCysLeu 160
Db 439 GCCTGCTGGCGCTGGCTCACCGCCAGCTGGCGCCCGCGCTGTGGCTGTCTGGCTG 498
QY 161 ProArgAlaSerHisArgPheValHisGlyThrAlaCysTyrAlaThrGlyTyrGlyAsp 180
Db 499 CCCCAGCGCTCACACCGCTTGGTGGCGGCACTGCTGGCGGCACTGCTGGCGGAGAC 558
QY 181 ValGlnGluAlaAspProLeuProLeuProTyrValLeuGlnGluValGluLeuArgLeu 200
Db 559 GTCCAGGAGGAGATCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 618
QY 201 LeuGlyGluAlaThrCysGlnCysLeuTyrSerGlnProGlyProPheAsnLeuThrLeu 220
Db 619 CTGGCGGAGGACCTGCTCAATGCTCTACAGCCAGCCGGTCCCTCAACCTCACTCTC 678
QY 221 GlnIleLeuProGlyMetLeuCysAlaGlyTyrProGluGlyArgArgAspThrCysGln 240
Db 679 CAGATATTGCCAGGAGATGCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 738
QY 241 GlyAspSerGlyGlyProLeuValCysGluGluGlyGlyValArgTyrPheGlnAlaGlyIle 260
Db 739 GGTGACTCTGGGCGGCGCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 798
QY 261 ThrSerPheGlyPheGlyCysGlyArgArgAsnArgProGlyValPheThrAlaValAla 280
Db 799 ACACGCTTTGGGTTGGCTGTGGACGAGAAACCGCCCTGGAGTTTCACTGTGTGGCT 858
QY 281 ThrTyrGluAlaTyrPheArgGluGlnValMetGlySerGluProGlyProAlaPhePro 300
Db 859 ACCTATGAGCATGATATGAGGAGGATGATGATGATGATGATGATGATGATGATGATG 918
QY 301 ThrGlnProGlnTyrThrGlnSerAspCysLeuHisGlnThrAlaPheLeuAspSerAla 320
Db 919 ACCAGCCCCAGAACCCAGTTCAGATTGTTTACATCAACGCGCATTCCTGGATTCTGCC 978
QY 321 ArgIleLeuLeuArgProLeuSerHisIleSerValGlyValSerThrGlyThrIleSer 340
Db 979 AGAATCCTTTGAGGCGCTTGTCCATATATATAGTAGGAGTCTCACTGGGACCAAAAGC 1038
QY 341 LeuValLeuProTyrLeuSerProHisSerLeuLeuGlyLeuTyrGlyPhe 357
Db 1039 CTTGTCTCTCCCTGGCTCTCTCCACACTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1089
RESULT 3
ABK31795
ID ABK31795 standard; DNA; 2457 BP.
XX
AC ABK31795;
XX
DT 23-APR-2002 (first entry)
XX
DE DNA encoding novel human protease #52.
XX Human; protease; cancer; immune-related disorder; cardiovascular disease;
KW neuronal-associated disease; metabolic disorder; inflammatory disorder;
KW nervous system disorder; sexual dysfunction; pain; mood disorder;
KW hypertension; psychotic disorder; neurological disorder; dyskinesia;
KW viral infection; human immunodeficiency virus; HIV; non-viral infection;
KW ocular disease; cytostatic; gene; ds.
XX
OS Homo sapiens.
XX
PN WO200200860-A2.

XX 03-JAN-2002.
XX
XX 26-JUN-2001; 2001WO-US020171.
XX
XX 26-JUN-2000; 2000US-0214047P.
XX
XX (SUGB-) SUGEN INC.
XX
XX Plowman G, Whyte D, Sudarsanam S, Manning G, Caenepeel S;
XX Charydzak G;
XX
XX WPI; 2002-139913/19.
XX
XX P-PSDB; AAU82753.
XX
XX Nucleic acids encoding novel human proteases, useful for useful for
XX treating diseases and disorders such as cancers, immune-related diseases
XX and disorders, cardiovascular disease (e.g. restenosis) and inflammatory
XX disorders.
XX
XX Claim 26; Fig 1SS-TT; 313pp; English.
XX
XX The present invention relates to the isolation of novel human proteases,
XX and the nucleic acids encoding them. The sequences of the invention are
XX useful for treating diseases and disorders such as cancers (e.g. breast,
XX colon, lung), immune-related diseases and disorders (e.g. inflammatory
XX diseases and asthma), cardiovascular diseases (e.g. restenosis and
XX coronary thrombosis), brain or neuronal-inflammatory disorders, metabolic
XX disorders (e.g. diabetes, obesity), central or peripheral nervous system
XX rheumatoid arthritis and psoriasis), mood disorders, attention
XX disorders, cognitive disorders, hypotension, hypertension, psychotic
XX disorders, neurological disorders (e.g. Alzheimer's disease, Parkinson's
XX disease) and dyskinesias. The nucleic acids and polypeptides are also
XX useful for treating viral infections caused by human immunodeficiency
XX virus (HIV), and non-viral infections such as ocular disease (e.g.
XX glaucoma) and macular degeneration. ABK31744-ABK31802 represent DNA
XX sequences encoding for the novel human proteases of the invention
XX
XX Sequence 2457 BP; 377 A; 851 C; 797 G; 432 T; 0 U; 0 Other;
SQ
Alignment Scores:
Pred. No.: 1-75e-74 Length: 2457
Score: 1505.00 Matches: 272
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 77.06% Indels: 0
DB: Gaps: 6
US-10-037-417-46 (1-357) x ABK31795 (1-2457)
QY 38 CysGlyArgProGluProSerAlaArgIleValGlyGlySerAsnAlaGlnProGlyThr 57
Db 112 TGGCGGCGGCTGAGCCCTCGGCCCGCTCGTGGGGGGCTCAACCGCGACGGGCGACC 171
QY 58 TrpProTyrGlnValSerLeuHisHisGlyGlyGlyHisIleCysGlyGlySerLeu 77
Db 172 TGGCCCTTGGCAAGTGGAGCTTGCACCTGAGAGTGGCCACATCTCGGGGGCTCCCTCATC 231
QY 78 AlaProSerTyrValLeuSerAlaAlaHisCysPheMetThrAsnGlyThrLeuGluPro 97
Db 232 GCCCTCTCTGGTCTCTCTCGCTGCTCACTGTTTCATGACGAATGGAGCGCTGGAGCCC 291
QY 98 AlaAlaGluTyrSerValLeuLeuGlyValHisSerGlnAspGlyProLeuAspGlyAla 117
Db 292 GCGCGCGAGTGTGGTACTGCTGGCGGCTGCATCTCCAGGACCGGCGCCCTGGACGGCGG 351
QY 118 HisThrArgAlaValAlaAlaIleValValProAlaAsnTyrSerGlnValGluLeuGly 137
Db 352 CACACCGCGGAGTGGCGGCGCATCTGTGTGCGGCGCACTACAGCAAGTGGAGCTGGGC 411
QY 138 AlaAspLeuAlaLeuArgLeuAlaSerProAlaSerLeuGlyProAlaValTyrPro 157

08-DEC-2000; 2000US-0254399P.
21-DEC-2000; 2000US-0257803P.
05-JAN-2001; 2001US-0260110P.
19-JAN-2001; 2001US-0262851P.
25-JAN-2001; 2001US-0264623P.
(INCYTE) INCYTE GENOMICS INC.

Yue H, Azimzai Y, Kalliock DA, Baughn MR, Griffin JA, Swarnakar A; Lal PG, Wallia NK, Hafalia A, Gandhi AR, Au-Young J, Elliott VS; Ramkumar J, Thangavelu K, Lu Y, Warren BA, Lu DAM, Lee EA; Tribouley CM, Arvizu C, Delegeane AM, Yao MG, Khan FA; Sanjanwala NM;
WPI; 2002-519664/55.
P-ESDB; ABB98135.

New isolated Protein Modification and Maintenance polypeptides, useful for diagnosis, and treatment of e.g. gastrointestinal disorders.

Claim 5; Page 196; 200pp; English.

The invention relates to an isolated Protein Modification and Maintenance (PMM) polypeptide. Polypeptides of the invention may be used in the diagnosis, treatment and prevention of disorders associated with decreased expression or activity of PMM. These include gastrointestinal disorders (e.g. Crohn's disease), cardiovascular disorders (e.g. anaemia), cell hypertension), autoimmune/inflammatory disorders, epithelial disorders proliferative disorders, developmental disorders (e.g. Alzheimer's disease) (e.g. scabies), neurological disorders (e.g. ectopic pregnancy), and in gene therapy or a vaccine for such diseases. They may also be used in the assessment of the effects of exogenous compound on the expression of nucleic acid and amino acid sequences of protein modification and maintenance molecules. The current sequence represents a human PMM encoding sequence of the invention, encoding a polypeptide which has been found to have homology CC with rat prolactin

XX Sequence 2681 BP; 431 A; 919 C; 861 G; 470 T; 0 U; 0 Other;

Alignment Scores:	1.9e-74	Length:	2681
Pred. No.:	1505.00	Matches:	272
Score:	100.00%	Conservative:	0
Percent Similarity:	100.00%	Mismatches:	0
Best Local Similarity:	77.06%	Indels:	0
Query Match:	6	Gaps:	0
DB:			

US-10-037-417-46 (1-357) x ABQ75956 (1-2681)

QY	38	CysGlyArgProGluProSerAlaAargileValGlyGlySerAsnAlaGlnProGlyThr	57
Db	112	TGCGGGGGCCCTGAGCCCTGGCCGCGCATCGTGGGGGGCTCAAAACGGCGACCGGGGAC	171
QY	58	TrpProTrpGlnValSerLeuHisGlyGlyGlyHisGlyHisGlyGlySerLeuLeu	77
Db	172	TGGCGCTTGGCAAGTGAGCCTGCACCATGAGGTGGCCACATCTGCGGGGGTCCCTCATC	231
QY	78	AlaProSerTrpValLeuSerAlaAlaHisCysPheMetThrAsnGlyThrLeuGluPro	97
Db	232	GCCCCCTCCTGGGTCTCTTCGCTGCTCACTGTTTCATGCAATGGACGCTGGAGGCC	291
QY	98	AlaAlaGluTrpSerValLeuGlyValHisSerGlnAspGlyProLeuAspGlyAla	117
Db	292	GCGCCGAGTGGTCGGTACTGCTGGCGGTGCACATCCACAGACGGGCCCTGGACGGCGCG	351
QY	118	HisThrArgAlaValAlaAlaAlaValProAlaAsnTyrSerGlnValGluLeuGly	137
Db	352	CACACCGCGCAGTGGCGGCATCGTGGTGGCGGCCACTACAGCCACTGGAGCTGGGC	411
QY	138	AlaAspLeuAlaLeuLeuArgLeuAlaSerProAlaSerLeuGlyProAlaValTrpPro	157

PN EPI130094-A2.
 XX 05-SEP-2001.
 XX 07-JUL-2000; 2000EP-00114089.
 XX 08-JUL-1999; 99JP-00194486.
 PR 11-JAN-2000; 2000JP-00118774.
 PR 02-MAY-2000; 2000JP-00183765.
 XX (HELI-) HELIX RES INST.
 XX Ota T, Nishikawa T, Isogai T, Hayashi K, Ishii S, Kawai Y;
 PI Wakamatsu A, Sugiyama T, Nagai K, Kojima S, Otsuki T, Koga H;
 DR WPI; 2001-524255/58.
 XX 830 Primers useful for synthesizing full length cDNA clones and their use
 PT in genetic manipulation.
 XX Claim 2; SEQ ID NO 650; 1380pp + Sequence Listing; English.
 CC The invention relates to primers for synthesizing full length cDNA
 CC clones. 830 cDNA molecules encoding a human protein have been isolated
 CC and nucleotide sequences of 5'- and 3'-ends of the cDNA molecules have
 CC been determined. Primers for synthesizing the full length cDNA are useful
 CC for clarifying the function of the protein encoded by the cDNA. The full
 CC length clones were obtained by construction of full length enriched cDNA
 CC libraries that were synthesised by the oligo-capping method. The primers
 CC enable the production of the full length cDNA easily without any special
 CC methods. The present sequence is the nucleotide sequence of the 5'-end of
 CC a cDNA provided in the invention. Note: The sequence data for this patent
 CC did not form part of the printed specification, but was obtained in CD-
 CC ROM format directly from EPO
 XX
 SQ Sequence 670 BP; 97 A; 245 C; 205 G; 118 T; 0 U; 5 Other;
 Alignment Scores:
 Pred. No.: 1.94e-39 Length: 670
 Score: 861.00 Matches: 167
 Percent Similarity: 94.92% Conservative: 1
 Best Local Similarity: 94.35% Mismatches: 5
 Query Match: 44.09% Indels: 4
 DB: 0 Gaps: 0
 US-10-037-417-46 (1-357) x AAK92190 (1-670)
 QY 38 CysGlyArgProGluProSerAlaArgIleVal-GlyGlySerAsnAlaGlnProGlyTh 57
 DB 143 TGGGGCGCCCTGAGCCCTCGCCCGCATCGTGGGGGGGCTCAACCGCGAGCGGGCAC 202
 QY 57 rTTPProTrpGlnValSerLeuHisHisGlyGlyHisIleCysGlyGlySerLeu11 77
 DB 203 CTGGCCCTGGCAAGTGGAGCCCTGCACCATGAGTGGCCATCATCTGCGGGGGCTCCCTCAT 262
 QY 77 eAlaProSerTrpValLeuSerAlaAlaHisCysPheMetThrAsnGlyThrLeuGluPr 97
 DB 263 CGCCCCCTCCCGGCTCTCTCCGTGCTCACTGTTTCATGACGAATGGACCGTGGAGCC 322
 QY 97 oAlaAlaGluTrpSerValLeuLeuGlyValHisSerGlnAspGlyProLeuAspGlyAl 117
 DB 323 CGGGCCGAGTGGTGGTACTGCTGGCGTCACTCCCGAGGACGGCCCTGGAGCGGC 382
 QY 117 aHisThrArgAlaValAlaAlaIleValValProAlaAsnTrpSerGlnValGluLeuGl 137
 DB 383 GCACACCGCGCAGTGGCGCCCATCGTGGTGGCGCCCACTACAGCCAGTGGAGCTGGG 442
 QY 137 yAlaAspLeuAlaLeuLeuArgLeuAlaSerProAlaSerLeuGlyProAlaValTrpPr 157
 DB 443 CGCCGACCTGGCCCTGCTGGCGCTGGCTCACCAGCCAGCTGGCGCCCGCGCTGGGCC 502
 QY 157 oValCysLeuProArgAlaSerHisArgPheValHisGlyThrAlaCysTrpAlaThrGl 177

Db 503 TGTCTGCCTGCCCGCCCTCAACCGCTTGTGACGGCAGCCGCTGTGGGCCACCGG 562
 QY 177 YTrpGlyAspVal-GlnGluAlaAspProLeuProLeuProLeuValLeuGln-GluVal 196
 Db 563 CTGGGGAGACNTCCANGGAGGCAATTTCTGTGCTCTCCCTGGTGTCTACAGGAAGTG 622
 QY 197 GluLeuArgLeuLeuGlyGlu-AlaThrCysGlnCysLeuTyr 210
 Db 623 GAGCTAAGTGTGCTGGCGAAGGCACTGCTCAATGTTCTCTAC 665
 RESULT 8
 AAK93456
 ID AAK93456 standard; cDNA; 670 BP.
 XX
 AC AAK93456;
 XX
 DT 06-NOV-2001 (first entry)
 XX Human cDNA clone representative sequence, SEQ ID NO: 1916.
 DE Human; full length cDNA; cDNA synthesis; oligo-capping; ss.
 XX Homo sapiens.
 OS
 PN EPI130094-A2.
 XX
 PD 05-SEP-2001.
 XX
 PF 07-JUL-2000; 2000EP-00114089.
 XX
 PR 08-JUL-1999; 99JP-00194486.
 PR 11-JAN-2000; 2000JP-00118774.
 PR 02-MAY-2000; 2000JP-00183765.
 XX (HELI-) HELIX RES INST.
 XX Ota T, Nishikawa T, Isogai T, Hayashi K, Ishii S, Kawai Y;
 PI Wakamatsu A, Sugiyama T, Nagai K, Kojima S, Otsuki T, Koga H;
 DR WPI; 2001-524255/58.
 XX 830 Primers useful for synthesizing full length cDNA clones and their use
 PT in genetic manipulation.
 XX Example 11; SEQ ID NO 1916; 1380pp + Sequence Listing; English.
 CC The invention relates to primers for synthesising full length cDNA
 CC clones. 830 cDNA molecules encoding a human protein have been isolated
 CC and nucleotide sequences of 5'- and 3'-ends of the cDNA molecules have
 CC been determined. Primers for synthesising the full length cDNA are useful
 CC for clarifying the function of the protein encoded by the cDNA. The full
 CC length clones were obtained by construction of full length enriched cDNA
 CC libraries that were synthesised by the oligo-capping method. The primers
 CC enable the production of the full length cDNA easily without any special
 CC methods. The present sequence was used as the representative sequence
 CC from a human clone which was used in homology searches to identify the
 CC clone. Note: The sequence data for this patent did not form part of the
 CC printed specification, but was obtained in CD-ROM format directly from
 CC EPO
 XX
 SQ Sequence 670 BP; 97 A; 245 C; 205 G; 118 T; 0 U; 5 Other;
 Alignment Scores:
 Pred. No.: 1.94e-39 Length: 670
 Score: 861.00 Matches: 167
 Percent Similarity: 94.92% Conservative: 1
 Best Local Similarity: 94.35% Mismatches: 5
 Query Match: 44.09% Indels: 4
 DB: 0 Gaps: 0
 US-10-037-417-46 (1-357) x AAK93456 (1-670)
 QY 38 CysGlyArgProGluProSerAlaArgIleVal-GlyGlySerAsnAlaGlnProGlyTh 57

which can be regulated for treating metastasis of malignant cells, inflammation, atherosclerosis, neurodegenerative disease and infections.

Disclosure; Fig 3; 11pp; English.

This invention comprises the cDNA and protein sequences of an isolated prostatic-like serine protease and reagents and methods for regulating the human prostatic-like enzyme activity. Prostatic is a trypsin-like serine protease purified from human seminal fluid. An antibody specific for prostatic-like serine protease is useful for immunodetection and diagnosis of micro-metastases, autoimmune lesions and renal failure in biopsy specimens, plasma samples and body fluids. The antibody may be used to modulate enzyme activity in a disease, such as metastasis of malignant cells, tumour angiogenesis, inflammation, chronic obstructive pulmonary disease (COPD), atherosclerosis, neurodegenerative disease (e.g. prion protein amyloid plaques of Genstmann-Straussler Syndrome, Creutzfeldt-Jakob disease, Scrapie) or infection, particularly viral infection. The human prostatic-like serine protease gene provides a therapeutic target of decreasing human prostatic-like serine protease activity, in particular for treating or preventing metastatic cancer. The agonists and antagonists of the nucleotide sequence may be used to mimic, augment and inhibit the enzyme activity which may be useful to treat osteoporosis, Paget's disease and degradation of bone implants particularly dental implants. Altered levels of human prostatic-like serine protease activity inhibits both smooth muscle cell proliferation and lipid accumulation and inhibit the progression of restenosis and atherosclerosis. The nucleic acid sequence is also useful in diagnostic assays for detecting diseases and abnormalities or susceptibility to diseases related to the presence of mutations in nucleic acid sequences which encode the enzyme. The present sequence represents the human prostatic-like serine protease #2 nucleotide sequence of the invention

Free. NO.:	739.00	Matches:	147
Score:	97.37%	Conservative:	1
Percent Similarity:	96.71%	Mismatches:	4
Best Local Similarity:	40.91%	Indels:	0
Query Match:	6	Gaps:	0
DB:			
32 10 037 417.45 (1-357) X ARK13566 (1-456)			

117-16 (1-357) y ARK13566 (1-456)

116

456	CCGCGCGCCAGTGGTCGGTACTGTGGCGGCGNTGCACCTCAAGACGCGGCCCTCTGGACGGC	397
Db	
Qy	117 AlaHisThrArgAlaValAlaAlaIleValValProAlaAsnTyrSerGlnValIleLeu	138
Db	396 GCGCACACCGCGCGAGTGGCGCCCATCTGTGGTGGCGGCGCAACTACAGCAACAGTGGAGCTG	333
Qy	137 GlyAlaAspLeuAlaLeuLeuArgLeuAlaSerProAlaSerLeuGlyProAlaValTyr	150
	
	277	

157 proVal[*cys*]enproArgAlaSerHisArgPheValHisGlyThrAlaCysTrpAlaThr 176

21

[illegible]

CCCTGCGCAGCCGCATCTGGTGTACAGGAAGTCAGGAAGTG 15'

216

97

22

[illegible]

DB
96 AACCTCACCTCCAGATATTGCCACCGGCAATGGTGTGTTT

QY 237 AspThrCysGlnGlyAspSerGlyGlyFluorenyl 2.10

1

|||||
36 GACACCTGCCAGGTGACTCTGGGGGGCCCTGGTTC 1

DB
RESULT 10
ID AAS69040
AC AAS69040 standard; cDNA; 597 BP.
XX
AC AAS69040;
XX
DT 13-FEB-2002 (first entry)
XX
DE DNA encoding novel human diagnostic protein #4844.
XX
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
OS Homo sapiens.
XX
PN WO200175067-A2.
XX
PD 11-OCT-2001.
XX
PF 30-MAR-2001; 2001WO-US008631.
XX
PR 31-MAR-2000; 2000US-00540217.
PR 23-AUG-2000; 2000US-00649167.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Drmanac RT, Liu C, Tang YT;
XX
DR WPI; 2001-639362/73.
DR P-PSDB; ABG04853.
XX
XX New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity.
XX
PS Claim 1; SEQ ID NO 4844; 103pp; English.
XX
XX The invention relates to isolated polynucleotide (I) and polypeptide (II)
CC sequences. (I) is useful as hybridisation probes, polymerase chain
CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
CC and in recombinant production of (II). The polynucleotides are also used
CC in diagnostics as expressed sequence tags for identifying expressed
CC genes. (I) is useful in gene therapy techniques to restore normal
CC activity of (II) or to treat disease states involving (II). (II) is
CC useful for generating antibodies against it, detecting or quantitating a
CC polypeptide in tissue, as molecular weight markers and as a food
CC supplement. (II) and its binding partners are useful in medical imaging
CC of sites expressing (II). (I) and (II) are useful for treating disorders
CC involving aberrant protein expression or biological activity. The
CC polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. AAS64197-AAS94564 represent novel human diagnostic
CC coding sequences of the invention. Note: The sequence data for this
CC patent did not appear in the printed specification, but was obtained in
CC electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 597 BP; 97 A; 217 C; 196 G; 87 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.: 2e-35 Length: 597
Score: 787.50 Matches: 150
Percent Similarity: 94.41% Conservative: 2
Best Local Similarity: 93.17% Mismatches: 7
Query Match: 40.32% Indels: 2
DB: 5 Gaps: 1

US-10-037-417-46 (1-357) x AAS69040 (1-597)
QY 93 GlyThrLeuGluProAlaAlaGluTrpSerValLeuLeuGlyVal-HisSerGlnAspG1 112
Db |||||
3 GGGACGCTGGAGCCGCGCGGCGAGTGGTCTGCTGGGGCGTCAACTCCAGGACGG 62
QY 112 yProLeuAspGlyAlaHisThrArgAlaValAlaAlaLeuValProAlaAsnTyrSe 132
Db |||||
63 GCGCTGGACGCGCGCACACCGCGCAGTGGCGCCATCGTGGTGGCGGCAACTACAG 122
QY 132 rGlaValGluLeuGlyAlaAspLeuAlaLeuArgLeuAlaSerProAlaSerLeuG1 152
Db |||||
123 CCAAGTGAGCTGGCGCGCGACCTGGCCCTGCTGGCCCTGCGCTCACCGCCAGGCTGG 182
QY 152 yProAlaValTrpProValCysLeuProArgAlaSerHisArgPheValHisGlyThrAl 172
Db |||||
183 CCGCGCGCTGGTGGCTGTCTGCTGCCCGCGGCTCACACCGCTTCGTGTCACGCGCCG 242
QY 172 aCysTrpAlaThrGlyTrpGlyAspValGlnGluAlaAspProLeuProTrpVa 192
Db |||||
243 CTGCTGGCGCACCGCTGGGAGAGCTCCAGAGGAGATCTCTGCCCTCTCCCTGGGT 302
QY 192 lLeuGlnGluValGluLeuArgLeuGlyGluAlaThrCysGlnCysLeuTyrSerG1 212
Db |||||
303 GCTACAGGAAGTGGAGCTAAGCTGCTGGCGGAGGCGCACCTGTCAATGTCTTACAGCCA 362
QY 212 nProGlyProPheAsnLeuThrLeuGlnIleLeuProGlyMetLeuCysAlaGlyTyrPr 232
Db |||||
363 GCGCGGTCCCTTCAACCTCACTCTCCAGATATGCGAGGATGCTGTGTGGCTACCC 422
QY 232 oGluGlyArgArgAspThrCysGln---GlyAspSerGlyGlyProLeuValCysGluG1 251
Db |||||
423 AGAGGCGCGCAGGACACCTGCCAGCCAGTCCAGAGCCCGGTCATGTGTCATCAGCCA 482
QY 251 u 251
Db 483 A 483
RESULT 11
ABT31936
ID APT31936 standard; DNA; 1733 BP.
XX
AC APT31936;
XX
DT 01-MAY-2003 (first entry)
XX
DE Human breast cancer / ovarian cancer related coding sequence #43.
XX
KW Human; gene; ds; cytostatic; breast cancer; ovarian cancer.
XX
OS Homo sapiens.
XX
PN WO2003000012-A2.
XX
PD 03-JAN-2003.
XX
PF 21-JUN-2002; 2002WO-US019773.
XX
PR 21-JUN-2001; 2001US-0300159P.
PR 27-JUN-2001; 2001US-0301351P.
XX
PA (MILL-) MILLENNIUM PHARM INC.
XX
PI Veiby OP;
XX
DR WPI; 2003-267848/26.
DR P-PSDB; ABJ37067.
XX
PT Determining the presence of breast cancer in an individual, involves
PT using specific polynucleotide markers.
XX
PS Disclosure; Page 192-193; 233pp; English.
XX

AAI98393 represent human KIM5C marker. Primer in the exemplification of the present invention

SQ Sequence 1796 BP; 327 A; 599 C; 511 G; 359 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.: 5.85e-30 Length: 1796
Score: 697.00 Matches: 160
Percent Similarity: 55.29% Conservative: 49
Best Local Similarity: 42.33% Mismatches: 133
Query Match: 35.69% Indels: 36
DB: 5 Gaps: 11

US-10-037-417-46 (1-357) x AAF98698 (1-1796)

```
QY 1 MetAlaGlnIysGlyValLeuGlyProGlyGlnLeuGlyAlaValAlaAenSerAspSer 20
Db 152 ATGCCCCAGAGGGGCTCTGGGGCTGGGAGCTGGGGCTGTGGCC----- 199
QY 21 TyrSerLeuTyr---GlyLeuValProSerGlyPro-----AlaArgGlyPro 35
Db 200 ATTCTGCTCTATCTTGGATTACTCCGGTCGGGACAGAGCGGAAGGGGAGAGCTCCC 259
QY 36 ProTyrCysGlyArgProGluProSerAlaArgIleValGlyGlySerAsnAlaGlnPro 55
Db 260 -----TGCAGT---GTGGCCCCCAAGCAGCATCACAGTGCACGATGCGATCGCC 310
QY 56 GlyThrTrpProTrpGlnValSerLeuHisGlyGlyGlyHisIleCysGlyGlySer 75
Db 311 GGTCACTGGCCCTGGCAGGTCAGATCACCTATGAGGCGGCCATGTGTGGTGGCTCT 370
QY 76 LeuIleAlaProSerTrpValLeuSerAlaAlaHisCysPheMetThrAsnGlyThrLeu 95
Db 371 CTCGTGCTGAGCAGTGGTCTGTCACTGCTCACTGCTTCCCCAGCAGCACCACAG 430
QY 96 GluProAlaAlaGluTrpSerValLeuLeuGlyValHisSerGlnAspGlyProLeuAsp 115
Db 431 GAA-----GCCTAGAGTCAAGCTGGGGGCCACAGTAGACTCTCTACTCCGAG 481
QY 116 GlyAlaHisThrArgAlaValAlaAlaIleValValProAlaAsnTyrSerGlnValGlu 135
Db 482 GACGCCAAGGTCAGCACCTGAAGGACATATCCCCACCCAGCTACTCTCCAGAGGGC 541
QY 136 LeuGlyAlaAspLeuAlaLeuArgLeuAlaSerProAlaSerLeuGlyProAlaVal 155
Db 542 TCCAGGGCGCATTTGCATCTCTCACTCAGCAGACCCATCCTCTCCCGCTACATC 601
QY 156 TrpProValCysLeuProArgAlaSerHisArgPheValHisGlyThrAlaCysTrpAla 175
Db 602 CGGCCCATCTGCTCCCTTCGAGCCAAACGCTTCCCCAGCGGCTCCACTGCAGCTGC 661
QY 176 ThrGlyTrpGlyAspValGlnGluAlaAspProLeuProLeuProTrpValLeuGlnGlu 195
Db 662 ACTGGCTGGGTCATGTGGCCCTCAGTGAGGCTCTCTGAGCCCAAGCCACTGCAGCA 721
QY 196 ValGluLeuArgLeuLeuGlyGluAlaThrCysGlnCysLeuTyrSerGlnProGlyPro 215
Db 722 CTCGAGGTGCCCTCATCAGTCGTGAGACGTGAATCTGCTGTACACATCGACGCCAAG 781
QY 216 PheAsnLeuThrLeuGlnIleLeuProGlyMetLeuCysAlaGlyTyrProGluGlyArg 235
Db 782 CTGAGGAGCCGCATTTGTCCAGAGGACATGGTGTGCTGGCTATGTGAGGGGGGC 841
QY 236 ArgAspThrCysGlnGlyAspSerGlyProLeuValCysGluGluGlyClyArgTrp 255
Db 842 AAGGAGCGCTGCAGGGTCACTCTGGGGGCCCACTCTCTGCTGTGGAGGCTCTCTGG 901
QY 256 PheGlnAlaGlyIleThrSerPheGlyPheGlyCysGlyArgArgAsnArgProGlyVal 275
Db 902 TACCTGACGGGATTTGTAGCTGGGAGATGCTGTGGGGCCCGCAACAGGCTGTGTG 961
QY 276 PheThrAlaValAlaThrTrpGluAlaTrpIleArgGluGlnValMetGlySerGluPro 295
Db 962 TACACTCTGGCTCCAGCTATGCTCTGTGATCCAAAGCAAGGTG-----ACAGAATC 1015
QY 296 GlyProAlaPheProThrGlnProGlnIlyThrGlnSerAsp-----CysLeuHis 312
```

```
Db 1016 CAGCCTCGTGTGTGTCCTCCCAACCCAGGAGTCCCGCCGAGCAACCTCTGTGGCAGC 1075
QY 313 GlnThrAlaPheLeuAspSer---AlaArgIleLeuLeuArgProLeuSerHisIleSer 331
Db 1076 CACCTGGGCTTTGAGCTCTGCCAGCCAGCGGCTTGCTGAGGCCATCTCTTCTGCTCT 1135
QY 332 ValGlyValSerThrGlyThrLysSerLeuValLeuProTrpLeuSer----- 347
Db 1136 CTGGGCTGGCTGTGGC-----CTCCTCTCCCATGGCTCAGCGAGCACTGAGCT 1186
QY 348 -----ProHisSerLeuLeuGlyLeuTrpGlyPhe 357
Db 1187 GGCCTTACTTCCAGGATGATGATCACAATCAAGACAGGAGCTTGCTCTTC 1240
RESULT 13
ABS76501
ID ABS76501 standard; cDNA; 1834 BP.
AC ABS76501;
XX
XX 11-DEC-2002 (first entry)
XX
XX cDNA encoding human ovarian cancer marker OV80.
XX Human; ovarian cancer; marker; cancer; familial history; brain disorder;
XX central nervous system disorder; bacterial meningitis; viral meningitis;
XX Alzheimer's disease; Parkinson's disease; cerebral oedema; hydrocephalus;
XX brain herniation; inflammation; encephalitis; testicular disorder;
XX nontuberculous granulomatous orchitis; connective tissue disorder;
XX heart disorder; ischaemic heart disease; atherosclerosis; neoplasm;
XX histological type; carcinogenic; ovarian cancer marker; gene; ss.
XX Homo sapiens.
XX
XX WO200271928-A2.
XX
XX 19-SEP-2002.
XX
XX 14-MAR-2002; 2002WO-US007826.
XX
XX 14-MAR-2001; 2001US-0276025P.
XX
XX 14-MAR-2001; 2001US-0276026P.
XX
XX 10-AUG-2001; 2001US-0311732P.
XX
XX 19-SEP-2001; 2001US-0323580P.
XX
XX 26-SEP-2001; 2001US-0324967P.
XX
XX 26-SEP-2001; 2001US-0325102P.
XX
XX 26-SEP-2001; 2001US-0325149P.
XX
XX (MILL-) MILLENNIUM PHARM INC.
XX
XX Monahan JB, Gannavarapu M, Hoersch S, Kamatkar S, Kovatis SG;
XX Meyers RE, Morrissey MF, Olandt PJ, Sen A, Vieby PO, Mills GB;
XX Bast RC, Lu K, Schmandt RE, Zhao X, Glatt K;
XX
XX WPI; 2002-723277/78.
XX
XX P-FSDB; ABG96402.
XX
XX Assessing whether a patient is afflicted with ovarian cancer, useful in
XX assessing the stage or progression of the disease, comprises comparing
XX the expression level of a cancer marker in a sample from a patient and
XX from a non cancer patient.
XX
XX Disclosure; Page 384; 481pp; English.
XX
XX The present invention relates to a new method for assessing whether a
XX patient is afflicted with ovarian cancer. The method involves comparing
XX the expression level of a marker in a patient sample and the normal level
XX of expression of the marker in a control non-ovarian cancer sample, where
XX the marker is selected from 363 cancer markers described in the
XX specification. The method of the invention is useful in diagnosing or
XX characterising cancer, in detecting the presence of cancer as early as
XX possible, and the recurrence of ovarian cancer. The method may also be of
```

[illegible][illegible]

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
84

Db 1264 GGCCCTACTTCCAGGATGCATCACACTCAAGGACAGAGCCTGTCCTTC 1317

RESULT 14

ABL67949

ID ABL67949 standard; DNA; 1834 BP.

XX

XX	
XX	
DT	15-MAY-2002 (first entry)
XX	
DE	Ovary cancer related gene sequence SEQ ID NO:6286.
XX	
KW	Human; cancer; colon; breast; ovary; oesophagus; kidney; thyroid;
KW	stomach; lung; prostate; pancreas; carcinoma; antitumour; cancerous;
XX	

KW cytochrome c gene; ds.
 KW
 XX
 XX
 OS Homo sapiens.
 XX
 XX
 PN WC200194629-A2.
 XX
 XX
 PD 13-DEC-2001.

XX	30-MAY-2001; 2001WO-US010838.
PF	
XX	
XX	
PR	05-JUN-2000; 2000US-0209473P.
PR	05-JUN-2000; 2000US-0209531P.
PR	18-SEP-2000; 2000US-0233133P.
PR	18-SEP-2000; 2000US-0233617P.
PR	20-SEP-2000; 2000US-0234009P.

PR 20-SEP-2000; 2000US-0234034P.
PR 20-SEP-2000; 2000US-0234052P.
PR 22-SEP-2000; 2000US-0234509P.
PR 22-SEP-2000; 2000US-0234567P.
PR 25-SEP-2000; 2000US-0234923P.
PR 25-SEP-2000; 2000US-0234924P.
PR 25-SEP-2000; 2000US-0235077P.
PR 25-SEP-2000; 2000US-0235083P.

PR 25-SEP-2000; 2000US-0235134P.
PR 25-SEP-2000; 2000US-0235280P.
PR 26-SEP-2000; 2000US-0235637P.

PR 26-SEP-2000; 2000US-0235638P.
 PR 27-SEP-2000; 2000US-0235711P.
 PR 27-SEP-2000; 2000US-0235720P.
 PR 27-SEP-2000; 2000US-0235840P.
 PR 27-SEP-2000; 2000US-0235863P.
 PR 28-SEP-2000; 2000US-0236028P.
 PR 28-SEP-2000; 2000US-0236032P.
 PR 28-SEP-2000; 2000US-0236033P.
 PR 28-SEP-2000; 2000US-0236034P.
 PR 28-SEP-2000; 2000US-0236109P.
 PR 28-SEP-2000; 2000US-0236111P.
 PR 29-SEP-2000; 2000US-0236842P.
 PR 29-SEP-2000; 2000US-0236891P.
 PR 02-OCT-2000; 2000US-0237172P.
 PR 02-OCT-2000; 2000US-0237173P.
 PR 02-OCT-2000; 2000US-0237278P.
 PR 02-OCT-2000; 2000US-0237294P.
 PR 02-OCT-2000; 2000US-0237295P.
 PR 02-OCT-2000; 2000US-0237316P.
 PR 03-OCT-2000; 2000US-0237425P.
 PR 03-OCT-2000; 2000US-0237598P.
 PR 03-OCT-2000; 2000US-0237604P.
 PR 03-OCT-2000; 2000US-0237606P.
 PR 03-OCT-2000; 2000US-0237608P.
 PR 01-NOV-2000; 2000US-0244867P.
 PR 01-NOV-2000; 2000US-0245084P.
 XX
 PA (AVAL-) AVALON PHARM.

XX Young PE, Augustus M, Carter KC, Ebner R, Endress G, Horrigan S;
 PI Soppet DR, Weaver Z;
 XX
 XX WPI; 2002-188264/24.

XX Screening for anti-neoplastic agent involves exposing cells to a chemical
 PT agent to be tested for anti-neoplastic activity, and determining a change
 PT in expression of a gene of a signature gene set.

XX Claim 1; SEQ ID NO 6286; 44pp; English.

XX The present invention describes a method (M1) for screening for an anti-
 CC neoplastic agent. The method involves exposing cells to a chemical agent
 CC to be tested for anti-neoplastic activity, determining a change in
 CC expression of at least one gene (I) of a signature gene set, where (I)
 CC comprises a sequence (S) selected from 8447 sequences (given in ABL61664
 CC to ABL70110), or is at least 95% identical to (S), where a change in
 CC expression is indicative of anti-neoplastic activity. (I) has cytostatic
 CC activity and can be used in gene therapy. M1 can be used for screening an
 CC anti-neoplastic agent, and can be used for producing a product which is
 CC the data collected with respect to the anti-neoplastic agent as a result
 CC of M1, and the data is sufficient to convey the chemical structure and/or
 CC properties of the agent. M1 can be used in the treatment of cancer such
 CC as colon, breast, stomach, lung, thyroid, oesophageal, ovarian, kidney,
 CC prostate or pancreatic cancer, adenocarcinoma, carcinoma, clear cell
 CC cancer, infiltrating ductal cancer, infiltrating lobular cancer, squamous
 CC cell carcinoma, neuroendocrine carcinoma, papillary carcinoma and Wilms
 CC tumour

XX Sequence 1834 BP; 309 A; 621 C; 526 G; 378 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.:	5,97e-30	Length:	1834
Score:	697.00	Matches:	160
Percent Similarity:	55.29%	Conservative:	49
Best Local Similarity:	42.33%	Mismatches:	133
Query Match:	35.69%	Indels:	36
DB:	6	Gaps:	11

US-10-037-417-46 (1-357) x ABL67949 (1-1834)

QY 1 MetAlaGlnLysGlyValLeuGlyProGlyGlnLeuGlyAlaValAlaAsnSerAspSer 20

DB 229 ATGGCCCAAGAGGGGCTCTGGGGCTTGGGAGCTGGGGGCTGTGGCC----- 276

QY	21	TyrSerLeuTyr---	GlyLeuValProSerGlyPro-----	AlaArgGlyPro 35
DB	277	ATTCTGCTCTATT	TGGATTACTCCGGTCGGGACAGGAGCGGAGAGGCGAAGCTCCC	336
QY	36	ProTyrCysGlyArgPro	GluProSerAlaArgIleValGlyGlySerAsnAlaGlnPro 55	
DB	337	-----TGC	GGT-----GTGGCCCCCAAGCAGCATACAGGTGGCAGGTCAGTCGCC 387	
QY	56	GlyThrTrpTrpGlnValSerLeuHis	GlyGlyHisIleCysGlySer 75	
DB	388	GGTCAGTGGCCCT	GGCAGTCCAGCATCACCTATGAAGCGCTCCATGTGTGGTGGCTCT 447	
QY	76	LeuIleAlaProSerTrpValLeuSerAla	IleHisCysPheMetThrAsnGlyThrLeu 95	
DB	448	CTCGTGTCTGAGCAGTGGGTGCTGT	CTCAGTCTCTCATCTTCCCCCAGGAGCAGCACCACAG 507	
QY	96	GluProAlaAlaGluTrpSerValLeuLeu	GlyValHisSerGlnAspGlyProLeuAsp 115	
DB	508	GAA-----	GCCTATGAGGTCAAGCTGGGGGCCACACAGTACACTTCTACTCCGAG 558	
QY	116	GlyAlaHisThrArgAlaValAlaAla	IleValProAlaAsnTyrSerGlnValGlu 135	
DB	559	GGAGCCAAAGGT	CAGCACCTCGAGGACATCATCCCCCAGCAGTACCTCCAGCAGGGC 618	
QY	136	LeuGlyAlaAspLeuAlaLeuLeuArgLeu	AlaSerProAlaSerLeuGlyProAlaVal 155	
DB	619	TCCCAGGGCGACATTCCTCCTCACT	CAGCAGACCCATCACCTTCTCCCGCTACATC 678	
QY	156	TrpProValCysLeuProArgAlaSerHis	ArgPheValHisGlyThrAlaCysTrpAla 175	
DB	679	CGGCCCCATCTGCTCCTCCCTGAC	CCAAACGGCTCTTCCCCAACAGCCACTGCAGTC 738	
QY	176	ThrGlyTrpGlyAspValGlnGluAlaAsp	ProLeuProLeuProTrpValLeuGlnGlu 195	
DB	739	ACTGGCTGGGGTCATGTGCCCCCT	CAGTAGAGCTCTCTCCCAACAGCCACTGCAGCAA 798	
QY	196	ValGluLeuArgLeuLeuGlyGluAlaThr	CysGlnCysLeuTyrSerGlnProGlyPro 215	
DB	799	CTCAGGTCCCTCTGATCAGTCTG	TGAGCGTAACTGCTGTACACATCGACGCCAAG 858	
QY	216	PheAsnLeuThrLeuGlnIleLeuProGly	MetLeuCysAlaGlyTyrProGluGlyArg 235	
DB	859	CTGAGGAGCCGACCTTCTCCAAAG	GAGCATGGTGTGTGGCTATGTGGAGGGGGC 918	
QY	236	ArgAspThrCysGlnGlyAspSerGly	GlyProLeuValCysGluGlyArgTrp 255	
DB	919	AAGGACGCTGCGCAGGGTGACTCT	GGGGGCCCACTCTCTGCGCTGTGGAGGGTCTCTGG 978	
QY	256	PheGlnAlaGlyIleThrSerPheGly	CysGlyArgAsnArgProGlyVal 275	
DB	979	TACCTGACGGGCATTGTGAGCTGG	GAGATGCTGTGGGGCCCGCAACAGGCTGTGTG 1038	
QY	276	PheThrAlaValAlaThrTyrGluAla	TrpIleArgGluGlnValMetGlySerGluPro 295	
DB	1039	TACACTGGCTCCAGCATGCTCTCT	CGATGCTCCAAAGCAAGGTG-----ACAGAACTC 1092	
QY	296	GlyProAlaPheProThrGlnProGln	LysThrGlnSerAsp-----CysLeuHis 312	
DB	1093	CAGCTCTGTGTGGTGGCCCCAAAC	CCAGGATCCCGCCAGCAGCAACCTCTGTGGCAGC 1152	
QY	313	GlnThrAlaPheLeuAspSer---	AlaArgIleLeuLeuArgProLeuSerHisIleSer 331	
DB	1153	CACCTGGCTTTCAGCTCTGCCCC	CCAGCGGCTTGTGTAGGGCCCATCTTTTCTCGCT 1212	
QY	332	ValGlyValSerThrGlyThrLysSer	LeuLeuProTrpLeuSer----- 347	
DB	1213	CTGGGCTGGCTCTGGGC-----	CTCCTCTCCCCATGGCTCAGCGAGCACTGAGCT 1263	
QY	348	-----	ProHisSerLeuLeuGlyLeuTrpGlyPhe 357	
DB	1264	GGCCCTACTTCCAGGATGGATGCAT	CACACTCAAGGACAGGAGCTGGTCTCTTC 1317	

RESULT 15
ABL68512
ID ABL68512 standard; DNA; 1834 BP.

XX AC ABL68512;
XX
XX
DT 15-MAY-2002 (first entry)
XX
DE Kidney cancer related gene sequence SEQ ID NO:6849.
XX
XX Human; cancer; colon; breast; ovary; oesophagus; kidney; thyroid;
XX stomach; lung; prostate; pancreas; carcinoma; antitumour; cancerous;
XX cytosolic; gene therapy; antineoplastic; Wilm's tumour; adenocarcinoma;
XX gene; ds.
XX
XX Homo sapiens.
XX
XX WO200194629-A2.
XX
XX 13-DEC-2001.
XX
XX 30-MAY-2001; 2001WO-US010838.
XX
XX 05-JUN-2000; 2000US-0209473P.
XX 05-JUN-2000; 2000US-0209531P.
XX 18-SEP-2000; 2000US-0233133P.
XX 18-SEP-2000; 2000US-0233617P.
XX 20-SEP-2000; 2000US-0234009P.
XX 20-SEP-2000; 2000US-0234034P.
XX 20-SEP-2000; 2000US-0234052P.
XX 22-SEP-2000; 2000US-0234509P.
XX 22-SEP-2000; 2000US-0234567P.
XX 25-SEP-2000; 2000US-0234923P.
XX 25-SEP-2000; 2000US-0234924P.
XX 25-SEP-2000; 2000US-0235077P.
XX 25-SEP-2000; 2000US-0235082P.
XX 25-SEP-2000; 2000US-0235134P.
XX 25-SEP-2000; 2000US-0235280P.
XX 26-SEP-2000; 2000US-0235637P.
XX 26-SEP-2000; 2000US-0235638P.
XX 27-SEP-2000; 2000US-0235711P.
XX 27-SEP-2000; 2000US-0235720P.
XX 27-SEP-2000; 2000US-0235840P.
XX 27-SEP-2000; 2000US-0235863P.
XX 28-SEP-2000; 2000US-0236028P.
XX 28-SEP-2000; 2000US-0236032P.
XX 28-SEP-2000; 2000US-0236033P.
XX 28-SEP-2000; 2000US-0236034P.
XX 28-SEP-2000; 2000US-0236109P.
XX 28-SEP-2000; 2000US-0236111P.
XX 29-SEP-2000; 2000US-0236842P.
XX 29-SEP-2000; 2000US-0236891P.
XX 02-OCT-2000; 2000US-0237172P.
XX 02-OCT-2000; 2000US-0237173P.
XX 02-OCT-2000; 2000US-0237278P.
XX 02-OCT-2000; 2000US-0237294P.
XX 02-OCT-2000; 2000US-0237295P.
XX 02-OCT-2000; 2000US-0237316P.
XX 03-OCT-2000; 2000US-0237425P.
XX 03-OCT-2000; 2000US-0237598P.
XX 03-OCT-2000; 2000US-0237604P.
XX 03-OCT-2000; 2000US-0237606P.
XX 03-OCT-2000; 2000US-0237608P.
XX 01-NOV-2000; 2000US-0244867P.
XX 01-NOV-2000; 2000US-0245084P.
XX
XX (AVAL-) AVALON PHARM.
XX
XX Young PE, Augustus M, Carter KC, Ebner R, Endress G, Horrikan S;
XX Soppet DR, Weaver Z;
XX
XX WPI; 2002-188264/24.
XX

PT Screening for anti-neoplastic agent involves exposing cells to a chemical
PT agent to be tested for anti-neoplastic activity, and determining a change
PT in expression of a gene of a signature gene set.

XX Claim 1; SEQ ID NO 6849; 44pp; English.

XX The present invention describes a method (M1) for screening for an anti-
CC neoplastic agent. The method involves exposing cells to a chemical agent
CC to be tested for anti-neoplastic activity, determining a change in
CC expression of at least one gene (I) of a signature gene set, where (I)
CC comprises a sequence (S) selected from 8447 sequences (given in ABL61664
CC to ABL70110), or is at least 95% identical to (S), where a change in
CC expression is indicative of anti-neoplastic activity. (I) has cytostatic
CC activity and can be used in gene therapy. M1 can be used for screening an
CC anti-neoplastic agent, and can be used for producing a product which is
CC the data collected with respect to the anti-neoplastic agent as a result
CC of M1, and the data is sufficient to convey the chemical structure and/or
CC properties of the agent. M1 can be used in the treatment of cancer such
CC as colon, breast, stomach, lung, thyroid, oesophageal, ovarian, kidney,
CC prostate or pancreatic cancer, adenocarcinoma, carcinoma, clear cell
CC cancer, infiltrating ductal cancer, infiltrating lobular cancer, squamous
CC cell carcinoma, neuroendocrine carcinoma, papillary carcinoma and Wilm's
CC tumour

XX Sequence 1834 BP; 309 A; 621 C; 526 G; 378 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.: 5 97e-30 Length: 1834
Score: 697.00 Matches: 160
Percent Similarity: 55.29% Conservative: 49
Best Local Similarity: 42.33% Mismatches: 133
Query Match: 35.69% Indels: 36
DB: 6 Gaps: 11

US-10-037-417-46 (1-357) x ABL68512 (1-1834)

Qy 1 MetAlaGlnIysGlyValLeuGlyProGlyGlnLeuGlyAlaValAlaAenSerAspSer 20
Db 229 ATG3CCCAAGAGGGGGTCTGGGGCTGGGAGCTGGGGCTGTGGCC----- 276
Qy 21 TyrSerLeuTyr---GlyLeuValProSerGlyPro-----AlaArgGlyPro 35
Db 277 ATTCTGCTCTATCTGTGGATTACTCCGGTCGGGGACAGAGCGGAGGGGAGAGCTCCC 336
Qy 36 ProTyrCysGlyArgProGluProSerAlaArgIleValGlyGlySerAsnAlaGlnPro 55
Db 337 -----TGGGGT---GTGGCCCCCAAGCAGCATCACAGGTGGCAGCAGTCAGTCGCC 387
Qy 56 GlyThrTrpProTrpGlnValSerLeuHisGlyGlyGlyHisIleCysGlyGlySer 75
Db 388 GGTGAGTGGCCCTGGCAGGTGAGTCACCTATGAAGCGCTCCATGTGTGTGGTGGCTCT 447
Qy 76 LeuIleAlaProSerTrpValLeuSerAlaAlaHisCysPheMetThrAsnGlyThrLeu 95
Db 448 CTCGTGTGTGAGCAGTGGGTGCTGTGTGCTGTCTGTCTGTCTGTCTGTCTGTCTGTCT 507
Qy 96 GluProAlaAlaGluTrpSerValLeuLeuGlyValHisSerGlnAspGlyProLeuAsp 115
Db 508 GAA-----GCCTATGAGGTCAAGCTGGGGGCCCCACCAGCTAGACTCTCTACTCCGAG 558
Qy 116 GlyAlaHisThrArgAlaValAlaAlaIleValValProAlaAsnTyrSerGlnValGlu 135
Db 559 GACGCCAAGGTTCAGCACCTCCAGGAGCATCATCCCCCAGCTACCTACCTCCAGAGGGGC 618
Qy 136 LeuGlyAlaAspLeuAlaLeuArgLeuAlaSerProAlaSerLeuGlyProAlaVal 155
Db 619 TCCAGGGGGGACATTGCACTCTCTCACTAGAGACCCCATCCTCTCTCCCGGTACATC 678
Qy 156 TrpProValCysLeuProArgAlaSerHisArgPheValHisGlyThrAlaCysTrpAla 175
Db 679 CGGCCCATCTCCCTCCCTGCAGCAACGCTCTTCCCAACGGCTCCCTCCACTGCACCTGTC 738
Qy 176 ThrGlyTrpGlyAspValGlnGluAlaAspProLeuProLeuProTrpValLeuGlnGlu 195

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: February 25, 2004, 15:57:04 ; Search time 2489 Seconds
(without alignments)
4283.165 Million cell updates/sec

Title: US-10-037-417-46
Perfect score: 1953
Sequence: 1 MAQKGVLPQLGAVANS...TSLVLPWLSPHSLGLWGWF 357

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 27513289 seqs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:

-MODEL=frame+p2n.model -DEV=xlp
-Q=/cgn2/1/USPTO_spool_p/US10037417/runat_24022004_101038_17460/app_query.fasta_1.519
-DB=EST -QPMT=fastap -SUFFIX=p2n.rst -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10037417@cgn 1 1 3549 @runat_24022004_101038_17460 -NCPU=6 -ICPU=3
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

- EST:
1: em_estba.*
2: em_esthum.*
3: em_estin.*
4: em_estmu.*
5: em_estov.*
6: em_estpl.*
7: em_estro.*
8: em_hic.*
9: gb_est1.*
10: gb_est2.*
11: gb_hic.*
12: gb_est3.*
13: gb_est4.*
14: gb_est5.*
15: em_estfun.*
16: em_estom.*
17: em_gas_hum.*
18: em_gas_inv.*
19: em_gas_pln.*
20: em_gas_vrt.*
21: em_gas_fun.*
22: em_gas_mam.*
23: em_gas_mus.*
24: em_gas_pro.*
25: em_gas_rod.*
26: em_gas_phg.*
27: em_gas_vrl.*
28: gb_gss1.*

29: gb_gss2.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
C 1	965	49.4	537	10	AW450407	UI-H-BI3-
C 2	799	40.9	456	9	AI190509	q449f07.x
C 3	613	31.4	1699	11	AK078696	Mus muscu
C 4	598.5	30.6	1629	11	AK010640	Mus muscu
C 5	589	30.2	1071	9	AL551470	AL551470
C 6	586.5	30.0	911	13	BX844660	BX844660
C 7	573.5	29.4	1150	11	AK080281	Mus muscu
C 8	565	28.9	1044	14	CD510038	CD510038
C 9	559.5	28.6	819	14	CA588481	hab64c07.
C 10	555.5	28.4	1188	11	BC034294	Homo sapi
C 11	551	28.2	938	14	CF238383	AGENCOURT
C 12	550	28.2	777	13	BX860661	BX860661
C 13	540.5	27.7	3055	11	AK004939	Mus muscu
C 14	540	27.6	2487	29	AY419856	Homo sapi
C 15	533	27.3	905	13	BQ729181	AGENCOURT
C 16	531	27.2	1050	11	AK006271	Mus muscu
C 17	530.5	27.2	1323	11	AK014645	Mus muscu
C 18	530	27.1	786	13	BX644781	BX644781
C 19	529.5	27.1	946	9	AL578261	AL578261
C 20	528.5	27.1	943	13	BQ222964	AGENCOURT
C 21	527.5	27.0	925	13	BX777626	BX777626
C 22	523.5	26.8	794	14	CF151097	AGENCOURT
C 23	523	26.8	2348	29	AY419858	Mus muscu
C 24	523	26.8	3248	11	BC008514	Mus muscu
C 25	523	26.8	4017	11	AK052738	Mus muscu
C 26	522.5	26.8	908	13	BQ946413	AGENCOURT
C 27	521	26.7	1010	13	BY710051	BY710051
C 28	517	26.5	947	9	AL555870	AL555870
C 29	517	26.5	1013	12	BM923713	AGENCOURT
C 30	516	26.4	784	14	CK029454	AGENCOURT
C 31	516	26.4	893	14	CF239681	AGENCOURT
C 32	514.5	26.3	969	14	CA488785	AGENCOURT
C 33	514	26.3	765	14	CA384091	AGENCOURT
C 34	512	26.2	889	12	BI218460	NC
C 35	510.5	26.1	2335	11	AK045663	Mus muscu
C 36	509.5	26.1	861	12	BI763558	603050292
C 37	506	25.9	744	14	CF998562	AGENCOURT
C 38	506	25.9	1201	13	BX417595	BX417595
C 39	505.5	25.9	802	12	BG962187	BG962187
C 40	505	25.9	881	14	CF242120	AGENCOURT
C 41	504.5	25.8	915	13	BUS24192	AGENCOURT
C 42	502.5	25.7	773	14	CK000011	AGENCOURT
C 43	498.5	25.5	728	12	BG969117	BG969117
C 44	498.5	25.5	1172	14	CD508404	CD508404
C 45	495	25.3	740	10	AW202972	fj24904.y

ALIGNMENTS

RESULT 1
AW450407/c
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE

AW450407
UI-H-BI3-akn-g-11-0-UI.s1 NCI CGAP_Sub5 Homo sapiens cDNA clone
IMAGE:2735037 3', mRNA sequence.
AW450407
AW450407.1 GI:6991183
EST.
Homo sapiens (human)
Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 537)


```

DEFINITION AL551470 Homo sapiens PLACENTA COT 25-NORMALIZED Homo sapiens cDNA
            clone CSODI064YF02 5-PRIME, mRNA sequence.
ACCESSION  AL551470
VERSION     AL551470.2 GI:31273286
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 1071)
            Li, W.B., Gruber, C., Jessee, J. and Polayes, D.
            Full-length cDNA libraries and normalization
            Unpublished (2001)
            On Feb 15, 2001 this sequence version replaced gi:12889449.
            Contact: Genoscope
            Genoscope - Centre National de Sequencage
            BP 191 91006 Evry cedex - France
            Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
            Library was constructed by Life Technologies, a division of
            Invitrogen. This sequence belongs to sequence cluster 7896.r For
            more information about this cluster, see
            http://www.genoscope.cns.fr/
            cgi-bin/cluster.cgi?seq=CSODI064DC01QP1&cluster=7896.r. Contact :
            Rong Liang Email : fliang@lifetech.com URL :
            http://fulllength.invitrogen.com/ Invitrogen Corporation 1600
            Faraday Avenue Genoscope sequence ID : CSODI064DC01QP1.
FEATURES    Location/Qualifiers
            source          1..1071
                        /organism="Homo sapiens"
                        /mol_type="mRNA"
                        /db_xref="taxon:9606"
                        /clone="CSODI064YF02"
                        /tissue_type="PLACENTA"
                        /clone_lib="Homo sapiens PLACENTA COT 25-NORMALIZED"
                        /note="First strand cDNA was primed with a NotI-oligo(dT)
                        primer. Five prime end enriched, double-strand cDNA was
                        digested with Not I and cloned into the Not I and EcoK v
                        sites of the pCMVSPORT 6 vector. Library was normalized."
ORIGIN
Alignment Scores:
Pred. No.:      2,81e-36      Length:      1071
Score:          589.00      Matches:      131
Percent Similarity: 57.73%      Conservative: 37
Best local Similarity: 45.02%      Mismatches:  108
Query Match:    30.16%      Indels:      16
DB:             Gaps:        6

US-10-037-417-46 (1-357) x AL551470 (1-1071)

Qy      1 MetAlaGlnlySgLyValLeuGlyProGlyGlnLeuGlyAlaValAlaAsnSerAspSer 20
Db      148 ATGCCCAAGAGGGGGTCTGGGGCTGGGCGAGCTGGGGCTGGGCC-----195
Qy      21 TyrSerLeuTyr---GlyLeuValProSerSeryPro-----AlaArgGlyPro 35
Db      196 ATTCGTCTCTATCTGTGATTACTCCGTGCGGGGACAGAGCGGAGCGGAGCTCCC 255
Qy      36 ProTyrCysGlyArgProGluProSerAlaArgIleValGlyGlySerAsnAlaGlnPro 55
Db      256 -----TGGCGT-----GTGGCCCCCAAGCAGCATCAGTGGCAGCATGGTGGCGCC 306
Qy      56 GlyThrTrpProTyrGlnValSerLeuHisGlyGlyGlyHisIleCysGlyGlySer 75
Db      307 GGTCAATGGCCCTGGCAGTCCAGCATCACCCTATGAAGCGTCCATGTGTGTGGTGTCT 366
Qy      76 LeuIleAlaProSerTrpValLeuSerAlaAlaHisCysPheMetThrAsnGlyThrLeu 95
Db      367 CTCGTGTCAGCAGTGGGTCTGTGACCTCTCACTGCTTCCCACGCGAGCACCACAAG 426
Qy      96 GluProAlaAlaGluTrpSerValLeuLeuGlyValHisSerGlnAspGlyProLeuAsp 115
Db      427 GAA-----GCTATGAGGTCAAGCTGGGGGCCACACGAGTAGACTCTCTCCGAG 477

```

```

Qy      116 GlyAlaHisThrArgAlaValAlaAlaIleValValProAlaAlaAsnTyrSerGlnValGlu 135
Db      478 GAGCCCAAGGTGAGCACCCTGAGGACATCATCCCCACCCAGCTACCTCCAGAGGGGC 537
Qy      136 LeuGlyAlaAspLeuAlaLeuLeuArgLeuAlaSerProAlaSerLeuGlyProAlaVal 155
Db      538 TCCAGGGGCGACATGTGCTCTCCCACTCAGCAGACCCATCACTTCTCCCGCTACATC 597
Qy      156 TrpProValCysLeuProArgAlaSerHisArgPheValHisGlyThrAlaCysTrpAla 175
Db      598 CGGCCCATCTGCTCTCTCGAGCAACGGCTCTCCCAACAGCGCTCCACTGCATCTGC 657
Qy      176 ThrGlyTrpGlyAspValGlnGluAlaAspProLeuProLeuProTyrValLeuGlnGlu 195
Db      658 ACTGGCTGGGGTCTATGTGGC-CCCTCAGTGAGCTCTCTGACGCCCAAGCCACTGCAGCA 716
Qy      196 ValGluLeuArgLeuLeuGlyGluAlaThrCysGlnCysLeuTyrSerGlnProGlyPro 215
Db      717 CTCGAGGTGCTCTGATCAGTCTGAGAGTGTAACTGCTGTACACATCGACGCCAAG 776
Qy      216 PheAsnLeuThrLeuGlnIleLeuProGlyMetLeuCysAlaGlyTyrProGluGlyArg 235
Db      777 CTGAGAGCGCGCACCTTTGTCCAAGAGGACATGTGTGTCTGTGCTATGTGGAGGGGGC 836
Qy      236 ArgAspThrCysGlnGlyAspSerGlyGlyProLeuValCysGlnGluGlyGlyArgTyr 255
Db      837 AAGACCGCTGCCAGGGTGACTCTGGGGGCCACTCTCTCGCTGTGGAGGGTCTCTGG 896
Qy      256 PheGlnAlaGlyIleThrSerPheGlyPheGlyCysGlyArgAsnArgProGlyVal 275
Db      897 TACCTGACGGGCATTTGTGAGCTGGGAGATGCTGTGGGGCCGCCAACAGCGCTGTGTG 956
Qy      276 PheThrAlaValAlaThrTyrGluAlaTrpIle 286
Db      957 TACACTGTGGCTCCAGCTAATGCTCTCTGATC 989

RESULT 6
BX844660      911 bp      mRNA      linear      EST 11-DEC-2003
LOCUS        BX844660 NICHDK_XGC_Emb4 Xenopus laevis cDNA clone IMAGE998C059551 ;
DEFINITION   IMAGE:4203196 5', mRNA sequence.
ACCESSION    BX844660
VERSION       BX844660.1 GI:39734664
KEYWORDS      EST.
SOURCE        Xenopus laevis (African clawed frog)
ORGANISM      Xenopus laevis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidae; Pipidae;
            Xenopodinae; Xenopus.
            1 (bases 1 to 911)
            Heil, O., Neubert, P., Peters, M., Radelof, U., Schneider, D.,
            Schroth, A., Korn, B. and Landgrebe, J.
            Xenopus laevis UniGene Set 1 (RZPDLIB No.988)
            Unpublished (2003)
            Contact: Ina Rolf
            RZPD Deutsches Ressourcenzentrum fuer Genomforschung GmbH
            Im Neuenheimer Feld 580, D-69120 Heidelberg, Germany
            RZPD: IMAGE998C059551.
            RZPDLIB: I.M.A.G.E. cDNA Clone Collection (amp- resistant) (RZPDLIB
            No.988) http://www.rzpd.de/cgi-
            bin/products/showlib.pl.cgi/response?libNo=998 RZPDLIB; Xenopus
            laevis UniGene Set 1 (RZPDLIB No.988) http://www.rzpd.de/cgi-
            bin/products/showlib.pl.cgi/response?libNo=988 Contact: Ina Rolf
            RZPD Deutsches Ressourcenzentrum fuer Genomforschung GmbH
            Heubnerweg 6, D-14059 Berlin, Germany
            Tel: +49 30 32639 101
            Fax: +49 30 32639 111
            www.rzpd.de
            This clone is available royalty-free from RZPD;
            contact RZPD (clone@rzpd.de) for further information. Seq primer:
            SP6, 5' ATTAGTGACACTATAG 3'.
            Location/Qualifiers
FEATURES

```

source

```

1. .911
/organism="Xenopus laevis"
/mol_type="mRNA"
/db_xref="taxon-8355"
/clone="IVAG998C059551 ; IMAGE:4203196"
/dev_stage="embryo, stage 31-32"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NICHD XGC Emb4"
/notes="Organ: whole embryo; Vector: pCMV-SPORT6; Site 1:
NotI; Site 2: SalI; Cloned unidirectionally. Primer: Oligo
dT. Average insert size 2.1 kb. Constructed by Life
Technologies. Note: This is a Xenopus Gene Collection
(XGC) library."

```

ORIGIN

Alignment Scores:

```

Pred. No.:      3,44e-36      Length:      911
Score:          586.50      Matches:      118
Percent Similarity: 57.09%      Conservative: 47
Best Local Similarity: 40.83%      Mismatches: 97
Query Match:      30.03%      Indels:      27
DB:              13          Gaps:        7

```

US-10-037-417-46 (1-357) x BX844660 (1-911)

```

QY 10 GlyGlnLeuGlyAlaValAlaAsnSerAspSerTyrSerLeuTyrGlyLeuValProSer 29
DB 71 GGGTCTCGAGGAGCATCATCATCAAGTTCTGAAGAACAATATGCTCTA-----121
QY 30 GlyProAlaArgGlyProProTyrCysGlyArgProGluProSerAlaArgIleValGly 49
DB 122 -----TCAGTCTGGGTGCGCAATGTTATCAGAACCGGATAGTAGG 163
QY 50 GlySerAsnAlaGlnProGlyThrTrpTrpGlnValSerLeuHisGlyGlyGly 69
DB 164 GGAACGACACATGAAGAGGGGAATGGCCCTGGCAGATCAGTCTGAGCTCAAAAGGGGAA 223
QY 70 HisIleCysGlyGlySerLeuIleAlaProSerTrpValLeuSerAlaAlaHisCysPhe 89
DB 224 CCTGCTCGGGGGATCCCTTATTGCCAACAGCTGGATTCTGCTGCTCTCACTGCTT 283
QY 90 MetThrAsnGlyThrLeuGluProAlaAlaGluTrpSerValLeuLeuGlyValHisSer 109
DB 284 GATTTCG-----CAAAACGTTTCCCAATATAAGATTACCTTGGGGTTAT---328
QY 110 GlnAspGlyProLeuAspGlyAlaHisThr-----ArgAlaValAlaAlaIleValVal 127
DB 329 CGATTGCTCTCTTCAAAGTCCCAACACAGTCTCCAGGAGTGTGAAGAAATCATCAT 388
QY 128 ProAlaAsnTyrSerGlnValGluLeuGlyAlaAspLeuAlaLeuLeuArgLeuAlaSer 147
DB 389 CACCCAGATTACCAATTTGAAGGATCTAATCGGACATTCGACTGATTGAATGGATCAA 448
QY 148 ProAlaSerLeuGlyProAlaValTrpProValCysLeuProArgAlaSerHisArgPhe 167
DB 449 CTTGTACCTTTACCCCTTACATCTCCAGCCCTCCCTCCACCCCGAGCTGCTGCTT 508
QY 168 ValHisGlyThrAlaCysTrpAlaThrGlyTrpGlyAspValGlnGluAlaAspProLeu 187
DB 509 CTGTCAGGGGTGAATGCTGGTGGTGGTGGTGGGAGATATCAAGAAGGACAGCCTCTA 568
QY 188 ProLeuProTrpValLeuGlnGlnValGluLeuArgLeuLeuGlyGluAlaThrCysGln 207
DB 569 GGTAAACCTAAACCTCGAAGAGCTACAGTGTAGTCTTAATTGACTGGCATAGCTGTGAG 628
QY 208 CysLeuTyrSer-----GlnProGly-----ProPheAsnLeuThrLeuGln 221
DB 629 TCCATGTACGAAACCCAGCTTGGGATACAAACCTAATGTTCTCTTT-----673
QY 222 IleLeuProGlyMetLeuCysAlaGlyTrpProGluGlyArgArgAspThrCysGlnGly 241
DB 674 ATCTCTGGATGACATGTTCTGTGCTGGCTACAGGAAGGAAGTAGACGCTGTGACGGA 733

```

```

QY 242 AspSerGlyGlyProLeuValCysGluGluGlyGlyArgTrpPheGlnAlaGlyIleThr 261
DB 734 GACTCTGGAGGACCTCTTGTATGTCAGAGTCAACAACACATGGTGGCAATACGCGATTGTC 793
QY 262 SerPheGlyPheGlyCysGlyArgArgAsnArgProGlyValPheThrAlaValAlaThr 281
DB 794 AGCTGGGGTATTGGATGTGGACAAAGTAAACACCGGGTGTCTATACCAAGTCCAGTAT 853
QY 282 TyrGluAlaIleTrpIleArgGluGlnVal 290
DB 854 TATGATGCTTGGATCAAGCAGTACATT 880

```

RESULT 7

AK080281

LOCUS

DEFINITION

Mus musculus 3 days neonate thymus cDNA, RIKEN full-length enriched

library, clone:A630023F11 product:similar to MARAPSN PRECURSOR (BC

3.4.21.-) [Homo sapiens], full insert sequence.

AK080281

VERSION

AK080281.1 GI:26348448

KEYWORDS

HTC; CAP trapper.

SOURCE

Mus musculus (house mouse)

ORGANISM

Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE

1

Carninci, P. and Hayashizaki, Y.

High-efficiency full-length cDNA cloning

Meth. Enzymol. 303, 19-44 (1999)

MEDLINE

99279253

PUBMED

10349636

REFERENCE

2

Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,

Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.

Normalization and subtraction of cap-trapper-selected cDNAs to

prepare full-length cDNA libraries for rapid discovery of new genes

Genome Res. 10 (10), 1617-1630 (2000)

MEDLINE

20499374

PUBMED

11042159

REFERENCE

3

Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P.,

Konno, H., Akiyama, J., Nishi, K., Kusunagi, T., Tashiro, H., Itoh, M.,

Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A.,

Yamanoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K.,

Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watabiki, M.,

Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsuura, S., Kawai, J.,

Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.

RIKEN integrated sequence analysis (RISA) system--384-format

sequencing pipeline with 384 multicapillary sequencer

Genome Res. 10 (11), 1757-1771 (2000)

MEDLINE

20530913

PUBMED

11076861

REFERENCE

4

The RIKEN Genome Exploration Research Group Phase II Team and the

FANTOM Consortium.

Functional annotation of a full-length mouse cDNA collection

Nature 409, 685-690 (2001)

JOURNAL

NATURE

409, 685-690 (2001)

REFERENCE

5

The FANTOM Consortium and the RIKEN Genome Exploration Research

Group Phase I & II Team.

Analysis of the mouse transcriptome based on functional annotation

of 60,770 full-length cDNAs

Nature 420, 563-573 (2002)

JOURNAL

NATURE

420, 563-573 (2002)

REFERENCE

6

(bases 1 to 1150)

Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Bono, H., Carninci, P.,

Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hashizume, W.,

Hayashida, K., Hayatsu, N., Hiramoto, K., Hiraoka, T., Hirozane, T.,

Hori, F., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T.,

Kato, H., Kawai, J., Kojima, Y., Kondo, S., Konno, H., Kouda, M.,

Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M.,

Nakamura, M., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Ohata, N.,

Okazaki, Y., Saito, R., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N.,

Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T.,

Soqabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Akahira, S.,
Takeda, Y., Tanaka, T., Tomaru, A., Toya, T., Yaeunishi, A.,
Muramatsu, M. and Hayashizaki, Y.

JOURNAL

Submitted (16-APR-2002) Yoshihide Hayashizaki, The Institute of
Physical and Chemical Research (RIKEN), Laboratory for Genome
Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
Kanagawa 230-0045, Japan (E-mail: genome-resgsc.riken.go.jp,
URL: http://genome.gsc.riken.go.jp/, Tel: 81-45-503-9222,
Fax: 81-45-503-9216)

CDNA library was prepared and sequenced in Mouse Genome
Encyclopedia Project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory in RIKEN.
Division of Experimental Animal Research in Riken contributed to
prepare mouse tissues.

Tissues were provided by Dr. John Todd (Dept. of Medical Genetics
Wellcome Trust Centre for Molecular Mechanisms in Disease Wellcome
Trust/MRC building Addenbrookes Hospital Cambridge) whose
assistance we gratefully acknowledge.

Please visit our web site for further details.

URL: http://genome.gsc.riken.go.jp/
URL: http://fantom.gsc.riken.go.jp/

FEATURES

source

Location/Qualifiers
1. 1150
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="FANTOM DB:A630023F11"
/db_xref="MGI:2406613"
/db_xref="taxon:10090"
/clone="A630023F11"
/tissue_type="thymus"
/clone_lib="RIKEN full-length enriched mouse cDNA library"
/dev_stage="3 days neonate"
66. 11052
/note="unnamed protein product; putative
similar to MARAPIN PRECURSOR (EC 3.4.21.-) [Homo sapiens]
(SWISSPROT|Q9BQR3, evidence: FASTV, 76.5%ID, 100%length,
match=876)";
/codon_start=1
/protein_id="BAC37864.1"
/db_xref="GI:26348449"
/translation="MRPHTAAILLLPILLSRGTEGARTILRACGHPFNRMYGGENA
LEGHPWQVSIQRNGIHFCGSLIAPTWLTAHCFNTSDISYQVLLGALKLQDPG
PHALYPYQVKSNPQVQMSADVALVELQGPVETNYLLPVCIPDPVIRESGMN
CWVTGSPSEQRDLRPNRVLKLAVEIIDTEKNLLYNKDVESDFOLTKIDMLCA
GFAGKKDACKGSGGLVCLVDQSVQAGVISWEGCGARRNRPVTVIRVISHKMT
QIIPELOFQGRAGTQQQKDSQGOORLAGNSAPCLAAHAWLALGALLLRIV"

CDS

Alignment Scores:
Pred. No.: 5.32e-35 Length: 1150
Score: 573.50 Matches: 133
Percent Similarity: 52.05% Conservative: 45
Best Local Similarity: 38.89% Mismatches: 133
Query Match: 29.37% Indels: 31
DB: 11 Gaps: 9

US-10-037-417-46 (1-357) x AK080281 (1-1150)

QY 26 LeuValProSerGlyPro-----AlaArgGlyProProTyrCysGlyArgProGluPro 43
Db 108 CTGCTCAGATCTGGAGACTGAGGGGCCAGGACTCTGAGAGCCTGTGGGCATCCAAAGATG 167
QY 44 SerAlaArgIleValGlyGlySerAsnAlaGlnProGlyThrTrpProTyrGlnValSer 63
Db 168 TTCACACCGATGTTGGCGGGGAGAACGCTTAGAAGGTGAGTGGCCCTGGCAGGTCAGC 227
QY 64 LeuHisHisGlyGlyGlyHisIleCysGlyGlySerLeuIleAlaProSerTrpValLeu 83
Db 228 ATCCAGCGCAGCGGAATCCACTTCTGTGTGGGGCAGCTCTCATCGCACCAACACATGGTCTCT 287

QY 84 SerAlaAlaHisCysPheMetThrAsnGlyThrLeuGluProAlaAlaGluTrpSerVal 103
Db 298 ACTGTCGACACTGCTTCCCAACTTCGGACATA-----TCCATATACCAAGGTC 338
QY 104 LeuLeuGlyValHisSerGlnAspGlyProLeuAspGlyAlaHisThrArgAlaValAla 123
Db 339 CTGCTGGGGGCACCTAAGCTGCAGCAGCA-----GGACCACACGCGCTGTATGTCCC 392
QY 124 AlalIleValValProAlaAsn-----TyrSerGlnValGluLeuGlyAlaAspLeuAla 141
Db 393 GTGAGCAGGTGAAGAAGCAATCTCAGTCAAGCATGGCTCCAGTCGCCACGCTGGT 452
QY 142 LeuLeuArgLeuAlaSerProAlaSerLeuGlyProAlaValTrpProValCysLeuPro 161
Db 453 CTGCTGGAGCTGCAGGGGCGGTGACCTTACCAATTCATCCTTCTGTGTCTTCCT 512
QY 162 ArgAlaSerHisArgPheValHisGlyThrAlaCysTrpAlaThrGlyTrpGlyAspVal 181
Db 513 GATCCCTCGGTCTATCTTGTAGTCGGGCATGAATGCTGGTCTCACTGGTGGGCGACCCC 572
QY 182 GlnGluAlaAspProLeuProTrpValLeuGlnGluValGluLeuArgLeuLeu 201
Db 573 AGTGAACAGACCGACTACCCCAACCCACGGTCTCTGCAGAACTTCTGTGCCCATCAIT 632
QY 202 GlyGluAlaThrCysGlnCysLeuTyrSerGlnProGlyProPheAsnLeuThrLeuGln 221
Db 633 GACACGCCCAAGTGCACACTGCTGTACAAAGATGTTGAGTCTGACTTCCAGCTCAA 692
QY 222 ---IleLeuProGlyMetLeuCysAlaGlyTyrProGluGlyArgArgAspThrCysGln 240
Db 693 ACCATCAAGATGACATGCTGTGCAGGCTTCGCAGAGGGGCAAGAGGACGCTGTAG 752
QY 241 GlyAspSerGlyGlyProLeuValCysGluGluGlyGlyArgTrpPheGlnAlaGlyIle 260
Db 753 GGTGACTCTGGAGGCCCATCTGGTGTCTGTGGACCATGCTGGTGGTGGAGGTGGGTG 812
QY 261 ThrSerPheGlyPheGlyCysGlyArgArgAsnArgProGlyValPheThrAlaValAla 280
Db 813 ATCAGCTGGAGAGAGGTGTGTGCCCGTCGGAATCGCCAGGTGTCTACATCCGTGTGACT 872
QY 281 ThrTyrGluAlaTrpIleArgGluGlnValMetGlySerGlu---ProGlyProAlaPhe 299
Db 873 TCTATCATCAAAATGGATCCCAAAATCATCCCAAGACTGCAGTTTCAGGGGAGGACGCG 932
QY 300 ProThrGlnProGlnLys-----ThrGlnSerAsp 309
Db 933 ACCCAGCAGCAACAAAAGACTCCCAAGGTCAGCAGCGCTGGCGGGAACCTCGCCCT 992
QY 310 CysLeuHisGlnThrAlaPheLeuAspSerAlaArgIleLeuLeuArgProLeuSerHis 329
Db 993 TGCCCTG-----GCTGCCCATGCATGTGTGTGGCTGGCCCTGGGAGCA 1031
QY 330 IleSerValGlyValSerThrGlyThrLysSerLeuValLeuProTrpLeuSerProHis 349
Db 1032 CTGCTGCTCAGATCGTCTAGGGCTGAGGGCTGCCAGCC-----CCACAGGTGGAGCCCC 1082
QY 350 SerLeu 351
Db 1083 CCACGTG 1088

RESULT 8

CD510038
LOCUS CD510038 1044 bp mRNA linear EST 12-JUN-2003
DEFINITION CDA99-E11.xid-t SHGC-CDA Gasterosteus aculeatus cdna clone
CDA99-E11 5', mRNA sequence.
CD510038
ACCESSION CD510038.1 GI:31440607
VERSION EST.
KEYWORDS Gasterosteus aculeatus (three spined stickleback)
SOURCE Gasterosteus aculeatus
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
Acanthomorpha; Acanthopterygii; Percomorpha; Gasterosteiformes;

Gasterosteidae; Gasterosteus.
 1. (Bases 1 to 1044)
 Kingsley, D.M., Peichel, C., Balabhadra, S., Grimwood, J., Dickinson, M., Schmutz, J., and Myers, R.M.
 Expressed sequence tags from *Gasterosteus aculeatus*
 Unpublished (2003)
 Contact: Kingsley, DM
 HMI and Department of Developmental Biology
 Stanford University School of Medicine
 Beckman Center B300, 279 Campus Drive, Stanford, CA 94305-5329, USA
 Tel: 650 725 5954
 Fax: 650 725 7739
 Email: kingsley@cmgm.stanford.edu
 Plate: 99
 High quality sequence stop: 793.
 Location/Qualifiers
 1. .1044
 /organism="Gasterosteus aculeatus"
 /mol_type="mRNA"
 /strain="Salinas river, CA"
 /db_xref="taxon:69293"
 /clone="CDA99-E11"
 /sex="mixed male and female"
 /tissue_type="heads and internal organs combined"
 /dev_stage="adult"
 /clone_lib="SHGC-CDA"
 /note="Vector: lambda ZAP Express/pBK-CMV; Site 1: EcoRI (5' adaptor); Site 2: XhoI (3' linker primer); The mixed organ cDNA library was generated using the ZAP-cDNA method by Stratagene. First strand cDNA synthesis was primed with a 50 bp linker primer containing an oligo dr sequence preceded by a synthetic XhoI site. 5 prime adaptors were used containing an EcoRI cohesive end. The finished cDNAs were inserted in to the ZAP express vector
 unidirectionally in the sense orientation with respect to the lacZ promoter of pBK-CMV. An amplified library was prepared from approximately 3 million primary clones in the lambda ZAP Express vector. In vivo excision was then used to generate individual pBK-CMV phagemid clones for EST sequencing."
 ORIGIN
 Alignment Scores:
 Pred. No.: 2.16e-34 Length: 1044
 Score: 565.00 Matches: 120
 Percent Similarity: 51.18% Conservative: 32
 Best Local Similarity: 40.40% Mismatches: 119
 Query Match: 28.93% Indels: 26
 DB: 14 Gaps: 4
 US-10-037-417-46 (1-357) x CD510038 (1-1044)
 QY 26 LeuValProSerGlyProAlaArgGlyProProtyrCysGlyArgProGluProSerAla 45
 DB 35 CTTCTCAGCAGGAGGAGCGACGACCGGATGTCGGCGCTGGACCGCGCAACAC 94
 QY 46 ArgIleValGlyGlySerAsnAlaGlnProGlyThrTrpProTrpGlnValSerLeuHis 65
 DB 95 CGCATCGTGGCGGGAGAACGGCCCGCGCTCCTGGCGGTGGAGTTCAGCTGCAC 154
 QY 66 HisGlyGlyGlyHisIleCysGlyGlySerLeuIleAlaProSerTrpValLeuSerAla 85
 DB 155 CGGTCTTCGGACACTCTGTGGGGGCTCGCTGTATCAACACCGCATGGGTCTCAGCGCC 214
 QY 86 AlaHisCysPheMetThrAsnGlyThrLeuGluProAlaAlaGluTrpSerValLeuLeu 105
 DB 215 GCGCACTGCTTCTCAGCACCACT-----CCTTCGGGTCTCAGCTGTACCTC 262
 QY 106 GlyValHisSerGlnAspGlyProLeuAspGlyAlaHisThrArgAlaValAlaLeu 125
 DB 263 GGCGGGACAGCCAGCAGCTGTCAACCTCAACGAGGGTCCCGCTCGTCTCTCTGCTC 322
 QY 126 ValValProAlaAsnTyrSerGlnValGluLeuGlyAlaAspLeuAlaLeuArgLeu 145

Db 323 ATCAGACATCCGGCTACAGCTCGCAGACCAACGACATCGCGCTCTTGGCGTG 382
 QY 146 AlaserProAlaserLeuGlyProAlaValTrpProValCysLeuProArgAlaserHis 165
 Db 383 TCCTCGCTGTGACCTTCAACACTACATCAGCCGGTGTGCTCGCGCAGAGGGAGC 442
 QY 166 ArgPheValHisGlyThrAlaCysTrpAlaThrGlyTrpGlyAspValGlnAlaAsp 185
 Db 443 GAGTTCTCGCGCGGAGCTCTCTGCTGCTCAGCGGTGGGGAACCATCGGTACAGCTT 502
 QY 186 ProLeuProLeuProTrpValLeuGlnGluValGluLeuArgLeuGlyGluAlaThr 205
 Db 503 CTTCTCCCGTTCCCTCAGCGCTTTCAGGAGGTTCAAGTCCCGCTGTGTCTCAACAGTCG 562
 QY 206 CysGlnCysLeuTyrSerGlnProGlyProPheAsnLeuThrLeuGlnIleLeuProGly 225
 Db 563 TGCACCGCCCTACGGA-----TCATACACGACCAAC 595
 QY 226 MetLeuCysAlaGlyTyrProGluGlyArgArgAspThrCysGlnGlyAspSerGlyGly 245
 Db 596 ATGCTCTGTCCGAGCGCGACGCGCGGGGGCGGACTCTCTGTCAGGAGAGCTCGGGGGC 655
 QY 246 ProLeuValCysGluGluGlyGlyArgTrpPheGlnAlaGlyIleThrSerPheGlyPhe 265
 Db 656 CCGCTGTGTAGTAAACAGGATCCAGGTGGTCCAGCGCGCGGTGGTGGCTTCGGGAAT 715
 QY 266 GlyCysGlyArgArgAsnArgProGlyValPheThrAlaValAlaThrTyrGluAlaTrp 285
 Db 716 CAGTGGCGGTGGCAGGGTTCCTCCGGGGTTCACGCGCGGTGTCCCGTACCACTGG 775
 QY 286 IleArgGluGlnVal-----MetGlySer 293
 Db 776 ATCAACAGCAGCGCGGAACCAACAGACGGGCTCGTGTCTTCATGGGGAACGGGACCA 835
 QY 294 GluProGlyProAlaPheProThrGlnProGlnLysThrGlnSerAspCys 310
 Db 836 CATTGGGAGCCCCG---CCGGCGCTCTCTGTTCGGGGGGCGTCTGCTGT 883
 RESULT 9
 CA588481 819 bp mRNA linear EST 19-NOV-2002
 LOCUS hab6c07.y1 Fugu U77 adult skin Takifugu rubripes cDNA clone
 DEFINITION IMAGE:6353269 5' similar to TR:Q9PVX7 Q9PVX7 EPIDERMIS SPECIFIC
 SERINE PROTEASE.; mRNA sequence.
 CA588481
 ACCESSION CA588481.1 GI:25133059
 VERSION EST
 KEYWORDS Takifugu rubripes (Fugu rubripes)
 SOURCE Takifugu rubripes
 ORGANISM Takifugu rubripes
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
 Acanthomorpha; Acanthopterygii; Petcomorpha; Tetraodontiformes;
 Tetraodontidae; Takifugu.
 1 (bases 1 to 819)
 Clark, M., Johnson, S.L., Lehrach, H., Lee, R., Li, F., Marra, M.,
 Eddy, S., Hillier, L., Kucaba, T., Martin, J., Beck, C., Wylie, T.,
 Underwood, K., Steptoe, M., Theising, B., Allen, M., Bowers, Y.,
 Person, B., Swaller, T., Gibbons, M., Pape, D., Harvey, N., Schurk, R.,
 Ritter, E., Kohn, S., Shin, T., Jackson, Y., Cardenas, M., McCann, R.,
 Waterston, R., and Wilson, R.
 WashU Zebrafish EST Project 1998
 Unpublished (1998)
 Contact: Stephen L. Johnson
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: zbrfish@wustl.edu
 Library materials provided by G. Elgar (UK MRC HGMP-RC) Library
 constructed by Drs. K. Kawakami, M. Sasaki, S. Sugano, K. Kikuchi
 and S. Watabe (University of Tokyo, Institute of Medical Science
 and Laboratory of Aquatic Molecular Biology and Biotechnology) DNA

Sequencing by: Washington University Genome Sequencing Center Clone
distribution: Fugu clone distribution information can be found
through the I.M.A.G.E. Consortium/LLNL, send email to:
info@image.llnl.gov
Seq primer: T3 ET from Amersham
High quality sequence stop: 490.

FEATURES

Location/Qualifiers
1..819
/organism="Takifugu rubripes"
/mol_type="mRNA"
/db_xref="taxon:31033"
/clone="IMAGE:6353269"
/sex="female"
/tissue type="skin"
/lab_host="DH10B (T1 phage-resistant)"
/clone_lib="Fugu U7 adult skin"
/notes="Vector: pBluescript-FL; Site 1: PFIM I (CCANNNTGG); Site 2: PFIM I (CCANNNTGG); BamHI-SmaI sites were converted to BamHI-PfIMI-SfiI-PfIMI sites (SmaI is destroyed). Other part of the vector is untouched. The cDNA is inserted between two PFIM I sites in T3 (5') to T7 (3') direction. Library materials provided by G. Elgar (UK MRC HMP-RC) and constructed and donated by Drs. K. Kawakami, M. Sasaki, S. Sugano, K. Kikuchi and S. Watabe (University of Tokyo, Institute of Medical Science and Laboratory of Aquatic Molecular Biology and Biotechnology)."

ORIGIN

Alignment Scores:
Pred. No.: 4.03e-34 Length: 819
Score: 559.50 Matches: 111
Percent Similarity: 56.52% Conservative: 32
Best Local Similarity: 43.87% Mismatches: 99
Query Match: 28.65% Indels: 11
DB: 14 Gaps: 3

US-10-037-417-46 (1-357) x CA588481 (1-819)

QY 38 CysGlyArgProGluProSerAlaAaGileValGlyGlySerAsnAlaGlnProGlyThr 57
DB 94 TGTGGAAAGACTCATGATCAACAGCAGATCTGTGGTGAGCAGGTGGCCCCCGAGGCGAGC 153
QY 58 TrpProTrpGlnValSerLeuHisHisGlyGlyHisLeCysGlyGlySerLeuLeu 77
DB 154 TGGCCCTGGCAGGCGCAGTCTGTGGGCTCACCGCTGTGGAGCTGTCTCATC 213
QY 78 AlaProSerTrpValLeuSerAlaAlaHisCysPheMetThrAsnGlyThrLeuGluPro 97
DB 214 AACAAACGATGGGTGTGAGCGCAGCCACTGTCTTCAG-----GGCGTCAGACCC 264
QY 98 AlaAlaGluTrpSerValLeuGlyValHisSerGlnAspGlyProLeuAspGlyAla 117
DB 265 AGCAGCGTG---ACGGTGTCTGGTGTGTCAGAGCCAGCGGCTCCACCCCAACGAG 321
QY 118 HisThrArgAlaValAlaAlaValValProAlaHisTrpSerGlnValGluLeuGly 137
DB 322 GAGCGGTGAGAGTACACAGATCATCAATCATCCGACTCAACTCAATCAATCAATCAAC 381
QY 138 AlaAspLeuAlaLeuGluLeuAlaSerProAlaSerLeuGlyProAlaValTrpPro 157
DB 382 AACGACATCTCCCTGTGCACTGGCAGAGCGGTGTCTTTTACAACTACATCCAGCCG 441
QY 158 ValCysLeuProArgAlaSerHisArgPheValHisGlyThrAlaCysTrpAlaThrGly 177
DB 442 GTCTGCTGGCGGCGGCGGAGACCTTCCACAGGCGCAGCGACTCTGTGGTCAACCGC 501
QY 178 TrpGlyAspValGlnGluAlaAspProLeuProLeuProTrpValLeuGlnGluValGlu 197
DB 502 TGGGCAACATGGATCGGAGTGTCTTCCCTTCCCTTCCCTTCCCTTCCCTTCCCTTCCCT 561
QY 198 LeuArgLeuLeuGlyGluAlaThrCysGlnCysLeuTySerGlnProGlyProPheAsn 217
DB 502 TGGGCAACATGGATCGGAGTGTCTTCCCTTCCCTTCCCTTCCCTTCCCTTCCCTTCCCT 561

DB 562 GTGCCCATCATGGGGGAACAGGGAGTGTAACTGTAACTACGGA----- 603
QY 218 LeuThrLeuGlnLeuLeuProGlyMetLeuCysAlaGlyTyrProGluGlyArgArgAsp 237
DB 604 ---GTGGGAAGGATTAACGACAACATGCTGTGTCGGGGTTCGGCAGCGGAGCAAGAT 660
QY 238 ThrCysGlnGlyAspSerGlyGlyProLeuValCysGluGluGlyGlyArgTrpPheGln 257
DB 661 TCTTGTCTCAGGGGGAATTCGGAGGCCCTTTGGTGATCAAGACAGACAGTCGCTGGATCCAG 720
QY 258 AlaGlyLeuThrSerPheGlyPheGlyCysGlyArgArgAsnArgProGlyValPheThr 277
DB 721 GCGGGGATCGTGAGTTTCGGCACCGGCTGTGTCACTACTACTACTACTACTACTACTACT 780
QY 278 AlaValAlaThrTyrGluAlaTrpIleArgGluGlnVal 290
DB 781 CGGGTCTCCAGTATAAAACCTGGATCAACAGTCATC 819

RESULT 10
BC034294

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

PUBLISHED

REFERENCE

AUTHORS

TITLE

JOURNAL

REMARK

COMMENT

BC034294 1188 bp mRNA linear HTC 17-DEC-2003
Homo sapiens cDNA clone IMAGE:4752547, containing frame-shift errors
BC034294
BC034294.1 GI:21706706
HTC.
Homo sapiens (human)
Homo sapiens
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 1188)
Strausberg, R.L., Feingold, E.A., Grouse, L.H., Derge, J.G., Klusner, R.D., Collins, F.S., Wagner, L., Shenmen, C.M., Schuler, G.D., Altshul, S.F., Zeeberg, B., Buetow, K.H., Schaefer, C.F., Bhat, N.K., Hopkins, R.F., Jordan, H., Moore, T., Max, S.I., Wang, J., Hsieh, F., Diatchenko, L., Marusina, K., Farmer, A.A., Rubin, G.M., Hong, L., Stapleton, M., Soares, M.B., Bonaldo, M.F., Casavant, T.L., Scheetz, T.E., Brownstein, M.J., Ustin, T.B., Toshnyuk, S., Carninci, P., Prange, C., Raja, S.S., Loquellano, N.A., Peters, G.J., Abramson, R.D., Miall, S.J., Bosak, S.A., McEwan, P.J., McKernan, K.J., Malek, J.A., Gunaratne, P.H., Richards, S., Wille, K.C., Hale, S., Garcia, A.M., Gay, L.J., Hulyk, S.W., Villalón, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A., Fahey, J., Hellon, E., Kettman, M., Madan, A., Rodriguez, S., Sanchez, A., Whiting, M., Madan, A., Young, A.C., Shevchenko, Y., Bouffard, G.G., Blakeley, R.W., Touchman, J.W., Green, E.D., Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M., Butlerfield, Y.S., Krzywinski, M.I., Skalska, U., Smal, M.A., Schnerch, A., Schein, J.E., Jones, S.J. and Marra, M.A.
Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences
Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
12477932
2 (bases 1 to 1188)
Strausberg, R.
Direct Submission
Submitted (02-JUL-2002) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA
NIH-MGC Project URL: <http://mgc.nci.nih.gov>
Contact: MGC help desk
Email: cgaps-r@mail.nih.gov
Tissue Procurement: James Cleaver, M.D.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Baylor College of Medicine Human Genome Sequencing Center
Center code: BCM-HGSC
Web site: <http://www.hgsc.bcm.tmc.edu/cdna/>
Contact: amg@bcm.tmc.edu
Gunaratne, P.H., Garcia, A.M., Lu, X., Hulyk, S.W., Louisse, H., Kowis, C.R., Sneed, A.J., Martin, R.G., Muzny, D.M., Nanavati,

A.N., Gibbs, R.A.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
 Series: IRAK Plate: 52 Row: d Column: 14
 This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA gi: 13994275
 This clone has the following problem: frame shifted.

FEATURES

Location/Qualifiers
 source
 1..1188
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:4752547"
 /tissue_type="Skin, squamous cell carcinoma"
 /clone_lib="NCI CGAP_Skn4"
 /lab_host="DH10B"
 /note="Vector: pCMV-SPORT6.1"

ORIGIN

Alignment Scores:
 Pred. No.: 1.5e-33 Length: 1188
 Score: 555.50 Matches: 128
 Percent Similarity: 50.79% Conservative: 32
 Best Local Similarity: 40.63% Mismatches: 118
 Query Match: 28.44% Indels: 37
 DB: 11 Gaps: 10

US-10-037-417-46 (1-357) x BC034294 (1-1188)

QY 32 AlaArgGlyProTyrCysGlyArgProGluProSerAlaArgIleValGlyGlySer 51
 Db 113 GCCAAGCAGCAACAGCCTGTGTGTCGCCAGGATGCTGAACCAATGTGTGGCGGCAG 172
 QY 52 AsnAlaGlnProGlyThrTrpProTyrGlnValSerLeuHisHisGlyGlyHisIle 71
 Db 173 GACACGAGGAGGCGAGTGGCCCTGGCACTGACATCCACGCCCAACGGAAGCCATTC 232
 QY 72 CysGlyGlySerLeuIleAlaProSerTrpValLeuSerAlaAlaHisCysPheMetThr 91
 Db 233 TGGGGGGCAGCCTCATGCGGAGCAGTGGTCTGACGGTGGCGACTGTTC----- 286
 QY 92 AsnGlyThrLeuGluProAlaAlaGluTrpSerValLeuLeuGlyValHisSerGlnAsp 111
 Db 287 CGCAACACCTCTGAGACGTCCCTG---TACCAGGTCTGTGTGGGGCAAGCAGCTAGTG 343
 QY 112 GlyProLeuAspGlyAlaHisThrArgAlaValAlaAlaIleValProAlaAsn--- 130
 Db 344 CAGCCG-----GGACCAACAGCTATGATGCCGGGTGAGCGAGGTGGAGCAACCCC 397
 QY 131 ---TyrSerGlnValGluLeuGlyAlaAspLeuAlaLeuAlaSerProAla 149
 Db 398 CTGTACCAGGCACGGCTCCAGCGCTGACCTGGCCCTGGTGGAGCTGGAGCACCAGTG 457
 QY 150 SerLeuGlyProAlaValTrpProValCysLeuProArgAlaSerHisArgPheValHis 169
 Db 458 CCCTTCCCAATTACATCTCCCGTGTGCTGCTGACCCCTCGGTGATCTTTGAGACG 517
 QY 170 GlyThrAlaCysTrpAlaThrClyTrpGlyAspValGlnGluAlaAspProLeuProLeu 189
 Db 518 GGCATGAACCTCTGGGTCACTGGTGGGGGACCCCCAGTGGAGGAAGACCTCTCCCGGAA 577
 QY 190 ProTrpValLeuGlnGluValGluLeuArgLeuLeuGlyGluAlaThrCysGlnCysLeu 209
 Db 578 CCGGGATCTCCAGAACTCGCTGTGCCCATCTCGACACACCACTGCAACCTGTCTC 637
 QY 210 TyrSerGlnProGlyProPheAsnLeuThrLeuGlnIleLeuProGly---MetLeuCys 228
 Db 638 TACAGCAAGACACCGAGTTTGGCTTACCAACCCAAACCAATCATCAATGATGTGTGC 697
 QY 229 AlaGlyTyrProGluGlyArgAspThrCysGlnGlyAspSerGlyGlyProLeuVal 248
 Db 698 GCCGCTTCGAGGGGCAAGAGATGCTGTCAAGGGCGAGCTCGGGCGGCCCTGTGTG 757

QY 249 CysGluGluGlyGlyArgTrpPheGlnAlaGlyIleThrSerPheGlyPheGlyCysGly 268
 Db 758 TGCTCGTGGTCACTGCTGCTGCGAGCGGGGTGATCAGCTGGGTGAGGGCTGTGCC 817
 QY 269 ArgArgAsnArgProGlyValPhe----- 276
 Db 818 CGCCAGAACCGCCAGGTGTCTATCATCCGTGTCAACCGGCCACCACTGGATCCATCG 877
 QY 277 -----ThrAlaValAlaThrTyrGluAlaTrpIleArgGluGlnValMetGly 292
 Db 878 GATCATCCCAAACTGCAGTCCAGCCAGCGAGGTTGGG---CGCCAGAGTG----- 928
 QY 293 SerGluProGlyProAlaPhePro-----ThrGlnProGlnLysThrGln 307
 Db 929 AGACCCCGGGGCGAGGAGCCCTTGAGCAGAGCTGTGCACCCAGCTGCCCGGCCA--- 985
 QY 308 SerAspCysLeuHisGlnThrAlaPheLeuAspSerAlaArgIle 322
 Db 986 -----CACCATCTGCTGTGTCCTCCACGAGCTGTGT 1018

RESULT 11

CF238383

LOCUS

AGENCOURT 15097797 NICHED XGC Emb6 Silurana tropicalis cDNA clone
 IMAGE:6993385 5', mRNA sequence.

ACCESSION

CF238383

VERSION

CF238383.1 GI:33441591

KEYWORDS

EST.

SOURCE

ORGANISM

Silurana tropicalis (western clawed frog)

Silurana tropicalis

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae; Pipidae;

Xenopodinae; Silurana.

1 (bases 1 to 938)

NIH-MGC <http://mgi.nci.nih.gov/>.

National Institutes of Health, Mammalian Gene Collection (MGC)

Unpublished (1999)

Contact: Daniela S. Gerhard, Ph.D.

Office of Cancer Genomics

National Cancer Institute / NIH

Bldg. 31 Rm10A07 Bethesda, MD 20892

Email: cgabs-x@mail.nih.gov

Tissue Procurement: Robert M. Grainger

cDNA Library Preparation: Life Technologies, Inc.

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Agencourt Bioscience Corporation

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

<http://image.llnl.gov>

Plate: LLAM14673 row: 1 column: 24

High quality sequence stop: 724.

FEATURES

Location/Qualifiers

1..938

/organism="Silurana tropicalis"

/mol_type="mRNA"

/db_xref="taxon:8364"

/clone="IMAGE:6993385"

/tissue_type="neurala"

/dev_stage="embryo, stages 14-19"

/lab_host="DH10B (phage-resistant)"

/clone_lib="NICHED XGC Emb6"

/note="Vector: pCMV-SPORT6.1; Site 1: NotI; Site 2: EcoRV;

Cloned unidirectionally. Primer: Oligo dt. Average insert

size 2.1 kb. Constructed by Invitrogen. Note: This is a

Xenopus Gene Collection (XGC) library."

ORIGIN

Alignment Scores:

Pred. No.: 2.35e-33 Length: 938

Score: 551.00 Matches: 113

Percent Similarity: 55.09% Conservative: 44

Best Local Similarity: 39.65% Mismatches: 103

Query Match:	28.21%	Indels:	25
DB:	14	Gaps:	7
US-10-037-417-46 (1-357) x CF238383 (1-938)			
Qy	38	CysGlyArgProGluProSerAlaAargIleValGlyGlySerAsnAlaGlnProGlyThr	57
Db	118	TGCGCGGTGCCAATAGTATCCAAATCGGATATAGGGGGAACGACAGTATGAAGGGGAA	177
Qy	58	TrpProTrpGlnValSerLeuHisHsIdgLyGlyHisIleCysGlyGlySerLeuIle	77
Db	178	TGGCCCTGGCAGATCAGCTCAGCTACAAAGGTCAAAGTCTCTGCGGAGGAAGCTTATC	237
Qy	78	AlaProSerTrpValLeuSerAlaAlaHisCysPheMetThrAsnGlyThrLeuGluPro	97
Db	238	ACCAGACAGCTGGGTCTCAGCTGCTCTCAGCTCTTGATGTCG-----CAAAAG	285
Qy	98	AlaAlaGluTrpSerValLeuLeuGlyValHisSerGlnAspGlyProLeuAspGlyAla	117
Db	286	GTTCCTCAATACATGTTTACCTGGAGTTAT---CAACTGTCTAACTTAAAAACCG	342
Qy	118	HisThr-----ArgAlaValAlaAlaIleValValProAlaAsnTyrSerGlnValGlu	135
Db	343	AATACGGTATCTAGTGGGGTGAAGAAATCATCATAAACAGGCTTACCAATACAGGGA	402
Qy	136	LeuGlyAlaAspLeuAlaLeuArgLeuAlaSerProAlaSerLeuGlyProAlaVal	155
Db	403	TCTAGCGGGGACATTTGCTCTGATTGAATCTGGAGAAACCGTCACCTTACCCTTACATC	462
Qy	156	TrpProValCysLeuProArgAlaSerHisArgPheValHisGlyThrAlaCysTrpAla	175
Db	463	CTCCAGTCTGCTCCCGCCCGGCTCTGAACCTTCTTCGACGGACCAAAATGCTGGGTG	522
Qy	176	ThrGlyTrpGlyAspValIleGlnIleAlaaspProLeuProLeuProTrpValLeuGlnGlu	195
Db	523	ACTGGATGGGAGATATCAAGAGAGCCCGGATCTAAGTAACCCGAAAAACCTCAGAG	582
Qy	196	ValGluLeuArgLeuLeuGlyGluAlaThrCysGlnCysLeuTyrSer-----	211
Db	583	GCTTCGGTGAAGCTAATGACTGGGAATAGCTGTGAGTCCATGTATGAACACCACTTCGGA	642
Qy	212	---GlnProGly---ProPheAsnLeuThrLeuGlnIleLeuProGlyMetLeuCysAla	229
Db	643	TACAAACCTAGTGTTCCTTT-----ATCATGACGACATGTTCTGTGCT	687
Qy	230	GlyTyrProGluGlyArgArgAspThrCysGlnGlyAspSerGlyGlyProLeuValCys	249
Db	688	GGATACAGGAAGGACAGATAGATGCTCTCAGGAGACTCTGGTGGGCTCTTGTATGC	747
Qy	250	GluGluGlyGlyArgTrpPheGlnAlaGlyIleThrSerPheGlyPheGlyCysGlyArg	269
Db	748	AATGTCAACAAACACCTGGTGGGAGTATGCATCATCAGCTGGGTATTGGATGTGCTGAA	807
Qy	270	ArgAsnArgPro-GlyValPheThrAlaValAlaThrTyrGluAlaTrpIleArgGluG	289
Db	808	GCCACGACACGGGGGTCTATACCAAGTCCAGTATTATGACTCTTGGATCAACAGTA	867
Qy	289	nValMetGlySerGluProGlyProAlaPheProThrGlnProGlnLysThrGlnSerAs	309
Db	868	CATT-----CCTTTCATTAATAATCAGCGAAGCAGGATGATGACTGG	909
Qy	309	pCysLeuHisGln	313
Db	910	TTGTGATACGAA	922
RESULT 12			
BX860661			
LOCUS			
DEFINITION			
BX860661 Rainbow trout multi-tissues (tcba) Oncochrynochus mykiss			
CDNA clone tcba009c.j.09 5prim, mRNA sequence.			
ACCESSION			
VERSION			
BX860661.1 GI:39958625			
KEYWORDS			
EST.			

Db 1660 ACAGAGGCAACCTCAGTGTACCGGAGGAGGAGCTGTAGCAGCGCTCAGATGAGAAG 1719

QY 44 -----SerAlaArgIleValGlySerAsn 52

Db 1720 GACTCGGAGTGTGGCTGGTTCATTCACGACAGCGCTGTGTGTGGGGGCGAGAT 1779

QY 53 AlaGlnProGlyThrTrpProTrpGlnValSerLeuHisHis---GlyGlyGlyHisIle 71

Db 1780 GCGGATAGGCGGAGTGGCCCTGGCAGGTAAAGCTGTCTGGGCCAGGCCACATC 1839

QY 72 CysGlyGlySerLeuIleAlaProSerTrpValLeuSerAlaAlaHisCysPheMetThr 91

Db 1840 TCGGCGTCTCCCTCATCTCCCAACGCTGGTCTCTGCCGACACATGCTACATCGAT 1899

QY 92 AsnGly-----ThrLeuGluProAlaAlaGluTrpSerValLeuLeuGlyValHisSer 109

Db 1900 GACGAGGATTGAGTACTACAGCCCGAGTGGAGCGCTTCTGGGCTTGCACGAC 1959

QY 110 Gln---AspGlyProLeuAspGlyAlaHisThrArgAlaValAlaAlaIleValValPro 128

Db 1960 CAGAGCCAGCGAGCGCCCTGGGTGAGGAGCGCGAGCTCAAGCGCATCTCTCCAC 2019

QY 129 AlaAsnTyrrSerGlnValGluLeuGlyAlaAspLeuAlaLeuArgLeuAlaSerPro 148

Db 2020 CCTCTTTCATGATGATTCACCTTCGATGATGATGATGATGATGATGATGATGAT 2079

QY 149 AlaSerLeuGlyProAlaValTrpProValCysLeuProArgAlaSerHisArgPheVal 168

Db 2080 CGAGAGTACAGCTCCATGTTGGGGCCCATCTCTGCGGAGCGCTCCATGTTCTTCCCT 2139

QY 169 HisGlyThrAlaCysTrpAlaThrGlyTrpGlyAspValGlnGluAlaAspProLeuPro 188

Db 2140 GCGCGCAAGGCCATCTGGGTCAAGCGCTGGGACACACACCATGATGAGGACCTGGCGG 2199

QY 189 LeuProTrpValLeuGlnGluValGluLeuArgLeuLeuGlyGluAlaThrCysGlnCys 208

Db 2200 CTG-----ATCCTGCAAAAGGAGTGGATCCCGCTCATCAACACGACACCTCGGAGAAC 2253

QY 209 LeuTyrrSerGlnProGlyProPheAsnLeuThrLeuGlnIleLeuProGlyMetLeuCys 228

Db 2254 CTCTGCGCGAG-----CAGATCACCGCGCGCATGATGTC 2289

QY 229 AlaGlyTyrrProGluGlyArgArgAspThrCysGlnGlyAspSerGlyGlyProLeu--- 247

Db 2290 GTGGCTTCTCCTCAGCGCGCGCTGAGACTCCTCCAGGCTGATTCGGGGGACCCCTGTCC 2349

QY 248 ValCysGluGluGlyGlyArgTrpPheGlnAlaGlyIleThrSerPheGlyPheGlyCys 267

Db 2350 AGCTGTGAGGCGGATGGCGGATCTTCAGGCGGTTGTTGAGTGGGAGACCGGCTGC 2409

QY 268 GlyArgArgAsnArgProGlyValPheThrAlaValAlaThrTyrrGluAlaTrpIleArg 287

Db 2410 GCTCAGAGGAACAGCGGCTGTACACAGGCTCCCTCTCTTTCGGGACTGGATCAAA 2469

QY 288 Glu 288

Db 2470 GAG 2472

RESULT 15

LOCUS BQ729181

DEFINITION AGENCOURT_8217490_NICHD_XGC_Emb2_Xenopus laevis cDNA clone

ACCESSION BQ729181

VERSION BQ729181.1

KEYWORDS BQ729181.1 GI:21868078

SOURCE EST.

ORGANISM Xenopus laevis (African clawed frog)

REFERENCE 1. (bases 1 to 905)

AUTHORS NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.

TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: Dr. Igor Dawid
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
DNA distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
Plate: LLAM10808 row: 1 column: 22
High quality sequence stop: 691.

FEATURES

source
1..905
Location/Qualifiers
/organism="Xenopus laevis"
/mol_type="mRNA"
/db_xref="taxon:8355"
/clone="IMAGE:4889637"
/tissue_type="embryo, stage 17/19"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NICHD XGC Emb2"
/note="Vector: pCMV-SPORT6; Site 1: NotI; Site 2: SalI;
Cloned unidirectionally. Primer: Oligo dt. Average insert
size 2.1 kb. Constructed by Life Technologies."

ORIGIN

Alignment Scores:
Pred. No.: 5,95e-32 Length: 905
Score: 533.00 Matches: 116
Percent Similarity: 55.90% Conservative: 45
Best Local Similarity: 40.28% Mismatches: 99
Query Match: 27.29% Indels: 29
DB: 13 Gaps: 7
US-10-037-417-46 (1-357) x BQ729181 (1-905)

QY 10 GlyGlnLeuGlyAlaValAlaAlaAsnSerAspSerTyrrSerLeuTyrrGlyLeuValProSer 29
Db 90 GGGTCTGAGGAGCAGCATCGCTATCAAGTCTCTGAGAAACATATGCTCTA-----140

QY 30 GlyProAlaArgGlyProProTyrrCysGlyArgProGluProSerAlaArgIleValGly 49
Db 141 -----TCAGTCTGCGGTGCGCCCAATGGTATCAGACGGATAGTAGG 182

QY 50 GlySerAsnAlaGlnProGlyThrTrpProTrpGlnValSerLeuHisGlyGly 69
Db 183 GGAACGGACAGTAAGAGGGGAATGGCCCTGGCAGATCAGTCTGAGCTCAAAAGGGAA 242

QY 70 HisIleCysGlyGlySerLeuIleAlaProSerTrpValLeuSerAlaAlaHisCysPhe 89
Db 243 CTGTCTGCGGGGATCCCTTATTGCCAACAGCTGGATTCGACTGCTGCTCCTGCTTT 302

QY 90 MetThrAsnGlyThrLeuGluProAlaAlaGluTrpSerValLeuGlyValHisSer 109
Db 303 GATTCG-----CAAAACGTTTCCCAATATAAGTTTACCTGGGGTTTAT---347

QY 110 GlnAspGlyProLeuAspGlyAlaHisThr-----ArgAlaValAlaAlaIleVal 127
Db 348 CGATGTCTCTCTCCATAAATCCCAACACAGCTCCAGAGGTGAAAGAAATCATATA 407

QY 128 ProAlaAsnTyrrSerGlnValGluLeuGlyAlaAspLeuAlaLeuArgLeuAlaSer 147
Db 408 CACCAGATTACCAATTGAAGGATCTAATGGGACATTCGACTGATTAATGATCA 467

QY 148 ProAlaSerLeuGlyProAlaValTrpProValCysLeuProArgAlaSerHisArgPhe 167
Db 468 CCGTGCACCTTTACCCCTTACATTCTCCAGCCCTCCACCCCGAGCTGCTGCTT 527

QY 168 ValHisGlyThrAlaCysTrpAlaThrGlyTrpGlyAspValGlnGluAlaAspProLeu 187
Db 528 CCGTGGGGTGAATGCTGGTGAATGCTGGTGAATGCTGGTGAATGCTGGTGAATGCTGCT 587

```
QY 188 ProLeuProTyrValLeuGlnGluValGluLeuArgLeuLeuGlyGluAlaThrCysGln 207
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
588 AGTAACCCCTAAACCCCTGCAGAGGCTACAGTGAGTCTAATTGACTGGCATAGCTGTGAG 647
QY 208 CysLeuTyrSer-----GlnProGly---ProPheAsnLeuThrLeuGln 221
Db ::||| ::||| ::||| ::||| ::||| ::||| ::||| ::|||
648 TCCATGTACGAAACCCAGCTTTGGGATACAAACCTAATGTCCTTTT----- 692
QY 222 IleLeuProGlyMetLeuCysAlaGlyTyrProGluGlyArgAspThrCysGlnGly 241
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
693 ATCCTGGATCACATGTTCTGTGGCTCAAGGAGGAAATATAGACGCTGTCTAGGGA 752
QY 242 AspSerGlyGlyProLeuValCysGluGluGlyGlyArgTyrPheGlnAlaGlyIleThr 261
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
753 GACTCTGGAGGACCTCTTGTATGCAGAGTCAACAACACATGTTGGCAATACGGCATTGT 812
QY 262 SerPhe-GlyPheGlyCysGlyArgAsnArgProGlyValPheThrAlaValAlaTh 281
Db ||||: ||| ||||| ||||: ||||: ||||: ||||: ||||: ||||: ||||:
813 AGCTGGGGGAATTGGATGTGGACAAGCTAACCAACCGGT-GGCTATACCAAGGTCCAGTA 871
QY 281 rTyrGluAlaTyrIleArgGlu 288
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
872 TTTGGAAGCTTGGATCAAGCAG 893
```

Search completed: February 25, 2004, 18:01:10
Job time : 2501 secs

150	SerLeuGlyProAlaValTrrProValCysLeuProArgAlaSerHisargPheValHis	161
Db	::: ::::	
438	CCCTTCACCAATTACATCTCTCCCGTGTGCTGCCCTGACCCCTCGTGATCTTTGAGACG	497
Qy	170 GlyThrAlaCysTrrAlaThrGlyTrrGlyAspValGlnGluAlaAspProLeuProLeu	189
Db	498 GCATGAACCTGCTGGGTCACTGGCTGGGGCAGCCCGAGTGAGGAACCTCTCTGCCCGAA	557
Qy	190 ProTrrValLeuGlnGluValGluAurGluLeuGlyGluAlaThrCysGlnCysLeu	209
Db	558 CGCGCGATCTCGCAAACTCGCTGTGCCCATCATCGACACACCCAAAGTGCACACCTGCTC	617
Qy	210 TyrSerGlnProGlyProPheAsnLeuThrLeuGlnIleLeuProGly	228
Db	618 TACAGCAAGACACCGAGTTTGGCTACCAACCCAAACCATCAAGATGACATGCTGTGC	677
Qy	229 AlaGlyTrrProGluGlyArgArgAspThrCysGlnGlyAspSerGlyGlyProLeuVal	248
Db	678 GCCTGCTTCGAGGAGGGCACAAGAGTGCCTGCAAGGGCGACTCGGGCGGCCCCCTGGTG	737
Qy	249 CysGluGluGlyGlyArgTrrPheGlnAlaGlyIleThrSerPheGlyPheGlyCysGly	268
Db	738 TGCTCGTGGGTCACTGTGCTCGAGCGGGGTGATCAGCTGGGGTGAGGGCTGTGCC	797
Qy	269 ArgArgAsnArgProGlyValPheThrAlaValAlaThrTrrGluAlaTrrIleArgGlu	288
Db	798 CGCCAGACCCGCCAGGTGTCTACATCCGTGTCAACCCGCCACCAACTGATCCATCGG	857
Qy	289 -----GlnValMetGlySerGlu-----Pro	295
Db	858 ATCATCCCCAAACTCAGTTCCA-GCCAGCGAGGTGGGGCGCCAGAAAGTGAGACCCCCG	916
Qy	296 GlyProAlaPhePro-----ThrGlnProGlnIlyThrGlnSerAspCys	310
Db	917 GGGCCAGAGCCCTTGAGCAGAGCTCTGCACCCAGCGCTGCCCGCCCA-----	964
Qy	311 LeuHisGlnThrAlaPheLeuAspSerAlaArgIle	322
	::: ::::	

RESULT 4
US-09-620-312D-431
; Sequence 431, Application US/09620312D
; Patent No. 6569662
; GENERAL INFORMATION:
; APPLICANT: Tang, Y. Tom
; APPLICANT: Liu, Chenghua
; APPLICANT: Asundi, Vinod
; APPLICANT: Zhang, Jie
; APPLICANT: Ren, Feiyan
; APPLICANT: Chen, Rui-hong
; APPLICANT: Zhao, Qing A.
; APPLICANT: Wehrman, Tom
; APPLICANT: Xue, Aidong J.
; APPLICANT: Yang, Yonghong
; APPLICANT: Wang, Jian-Rui
; APPLICANT: Zhou, Ping
; APPLICANT: Ma, Yunding
; APPLICANT: Wang, Dunrui
; APPLICANT: Wang, Zhiwei
; APPLICANT: John Tillinghast
; APPLICANT: Drmanac, Radoje T.
; TITLE OF INVENTION: No. 6569662el Nucleic Acids and
; TITLE OF INVENTION: Polypeptides
; FILE REFERENCE: 784CIP2B
; CURRENT APPLICATION NUMBER: US/09/620,312D
; CURRENT FILING DATE: 2000-07-19
; PRIOR APPLICATION NUMBER: 09/552,317
; PRIOR FILING DATE: 2000-04-25
; PRIOR APPLICATION NUMBER: 09/488,725
; PRIOR FILING DATE: 2000-01-21
; NUMBER OF SEQ ID NOS: 1105
; SOFTWARE: pt FL genes Version 1.0

```

; SEQ ID NO 431
; LENGTH: 1212
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (135)..(1007)
US-09-620-312D-431

Alignment Scores:
Pred. No.: 2,69e-37 Length: 1212
Score: 562.50 Matches: 128
Percent Similarity: 51.28% Conservatives: 32
Best Local Similarity: 41.03% Mismatches: 121
Query Match: 28.80% Indels: 32
DB: 4 Gaps: 9

US-10-037-417-46 (1-357) x US-09-620-312D-431 (1-1212)
QY 32 AlaArgGlyProProTyrCysGlyArgProGluProSerAlaAlaArgIleValGlyGlySer 51
Db 192 GCCAGGCGACACACCCCTGTGTGTCGCCACGAGTCTGAACCGATGTGTGGCGGGCAG 251
QY 52 AsnAlaGlnProGlyThrTrpProTrpGlnValSerLeuHisHisGlyGlyGlyHisIle 71
Db 252 GACACGAGGAGGCGAGTGGCCCTGGCAAGTCAGCATCCAGCCAGCAACGAGCCACTTC 311
QY 72 CysGlyGlySerLeuIleAlaProSerTrpValLeuSerAlaAlaHisCysPheMetThr 91
Db 312 TGGCGGGGCGAGCTCATCGCGAGCAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTTC 365
QY 92 AsnGlyThrLeuGluProAlaAlaGluTrpSerValLeuLeuGlyValHisSerGlnAsp 111
Db 366 CGCAACACCTCTGAGAGGTCCTG---TACCAGTCTCTGCGGGGCGAGGCGAGTGTG 422
QY 112 GlyProLeuAspGlyAlaHisThrArgAlaValAlaAlaIleValValProAlaAsn--- 130
Db 423 CAGCGG-----GGACACACGCTATGTATGTCGCCGGTGGAGCGAGTGGAGAGCAACCC 476
QY 131 ---TyrSerGlnValGluLeuGlyAlaAspLeuAlaIleLeuArgLeuAlaSerProAla 149
Db 477 CTGTACGAGGCGACGGCTCCAGCGCTGAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 536
QY 150 SerLeuGlyProAlaValTrpProValCysLeuProArgAlaSerHisArgPheValHis 169
Db 537 CCCTTACCAATACATCTCTCCCGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 596
QY 170 GlyThrAlaCysTrpAlaThrGlyTrpGlyAspValGlnGluAlaAlaAspProLeu 189
Db 597 GGCATGAACCTGCTGGTCACTGGCTGGGCGAGCCAGTGGAGAGACCTCTCTGCCCGAA 656
QY 190 ProTrpValLeuGlnGluValGluLeuArgLeuLeuGlyGluAlaThrCysGlnCysLeu 209
Db 657 CCGCGGATCTCGAGAACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 716
QY 210 TyrSerGlnProGlyProPheAsnLeuThrLeuGlnIleLeuProGly---MetLeuCys 228
Db 717 TACAGCAAGACACCGAGTTGGCTACCAACCCCAACCATCATGAATGACATGCTGTGTC 776
QY 229 AlaGlyTyrProGluGlyArgArgAspThrCysGlnGlyAspSerGlyGlyProLeuVal 248
Db 777 GCGGCTTCGAGGAGGCGCAAGAGATGCTGCAAGGCGGACTCGGCGCGCCCTCTGGTG 836
QY 249 CysGluGluGlyGlyArgTrpPheGlnAlaGlyIleThrSerPheGlyPheGlyCysGly 268
Db 837 TSCCTCTGGTTCAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 896
QY 269 ArgArgAsnArgProGlyValPheThrAlaValAlaThrTyrGluAlaTrpIleArgGlu 288
Db 897 CGCCAGAACCGCCAGGTTGTATCATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 956
QY 289 -----GlnValMetGlySerGlu-----Pro 295

```


Thu Feb 26 13:50:26 2004

```
RESULT 7
US-09-386-642-7
; Sequence 7, Application US/09386642
; Patent No. 6420157
; GENERAL INFORMATION:
; APPLICANT: Darrow, Andrew
; APPLICANT: Qi, Jensen
; APPLICANT: Andrade-Gordon, Patricia
; TITLE OF INVENTION: Zymogen Activation System
; FILE REFERENCE: ORT-1028
; CURRENT APPLICATION NUMBER: US/09/386,642
; CURRENT FILING DATE: 1999-08-31
; NUMBER OF SEQ ID NOS: 60
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 7
; LENGTH: 1169
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Fusion gene
; OTHER INFORMATION: with homo sapien serine protease catalytic domain
US-09-386-642-7

Alignment Scores:
Pred. No.: 2,72e-36 Length: 1169
Score: 550.00 Matches: 113
Percent Similarity: 56.20% Conservative: 41
Best Local Similarity: 41.24% Mismatches: 112
Query Match: 28.16% Indels: 8
DB: Gaps: 3

US-10-037-417-46 (1-357) x US-09-386-642-7 (1-1169)
QY 46 ArgileValGlyGlySerAsnAlaGlnProGlyThrTrpProGlnValSerLeuHis 65
Db 163 AAGATCGTTGGGGCTATGCTCTAGAGCGCGTCAAGTGGCCCTGGCAGGTCAGCATCACC 222
QY 66 HisGlyGlyHisLeuValGlyGlySerLeuLeuAlaProSerTrpValLeuSerAla 85
Db 223 TATGAAGCGCTCCATGTTGTGGTGTCTCTGTGTCTGAGCGTGGTGTCTGTAGCT 282
QY 86 AlaHisCysPheMetTrpAsnGlyThrLeuGluProAlaAlaGluTrpSerValLeu 105
Db 283 GCTCAGTCTTCCCGAGCGAGCACCACAGGAA-----GCTATGAGGTCAAGCTG 333
QY 106 GlyValHisSerGlnAspGlyProLeuAspGlyAlaHisThrArgAlaValAlaAla 125
Db 334 GGGGCCACCAAGTAGATCTCTACTCCGAGGAGCCCAAGGTCAGACCCCTGAAGGACATC 393
QY 126 ValValProAlaAsnTyrSerGlnValGluLeuGlyAlaAspLeuAlaLeuArgLeu 145
Db 394 ATCCCCACCCAGCTACTCTCCAGAGGGCTCCAGGGCGACATTCGACTCTCCAACTC 453
QY 146 AlaSerProAlaSerLeuGlyProAlaValTrpProValCysLeuProArgAlaSerHis 165
Db 454 AGCAGACCCATCACCTCTCCCGCTACATCCGGCCCATCTGCTCCCTCGACGCCAACGCC 513
QY 166 ArgPheValHisGlyThrAlaCysTrpAlaThrGlyTrpGlyAspValGlnGluAlaAsp 185
Db 514 TCCCTTCCCAACGGGCTCCACTGCACTGTGCTGGGGTCATGTGGCCCTTCAGTG 573
QY 186 ProLeuProLeuProTrpValLeuGlnGluValGluLeuArgLeuLeuGlyGluAlaThr 205
Db 574 AGCCTCTCAGCCCAAGCCATTCGAGCACTCGAGGTGCCTCTGATCAGTGTGAGAGC 633
QY 206 CysGlnCysLeuTyrSerGlnProGlyProPheAsnLeuThrLeuGlnLeuProGly 225
Db 634 TGTAACTGCTGTGACAACTACGACCGCAAGCCTGAGGAGCCGCACTTTTCTCCAGAGGAC 693
QY 226 MetLeuCysAlaGlyTyrProGluGlyArgArgAspThrCysGlnGlyAspSerGlyGly 245
Db 694 ATGGTGTGTGCTGTGCTATGTGGAGGGGGGCAAGGACGCGCTGCCAGGCTGACTCTGGGGGC 753
```

```
246 ProLeuValCysGluGluGlyGlyArgTrpPheGlnAlaGlyIleThrSerPheGlyPhe 265
Db 754 CCACTCTCTCTGCTGTGTGGAGGTCTCTGTGTACTGAGCGGCATTTGAGCTGGGAGAT 813
QY 266 GlyCysGlyArgArgAsnArgProGlyValPheThrAlaValAlaThrTrpGluAlaTrp 285
Db 814 GCCTGTGGGGCCCAACAGGCTTGGTGTGTACTCTTGGCTCCAGCTATGCTCTCTGG 873
QY 286 IleArgGluGlnValMetGlySerGluProGlyProAlaPheProThrGlnProGlnLys 305
Db 874 ATCCAAAGCAAGGTG-----ACAGAACTCCAGGCTCTGTGTGGTGGCCCAACCCAGGAG 927
QY 306 ThrGlnSerAsp-----CysLeuHisGlnThrAlaPhe 316
Db 928 TCCAGCGCCGACAGCAACTCTGTGGCAGCCACCTGGCCCTTC 969

RESULT 8
US-09-027-337-1
; Sequence 1, Application US/09027337B
; Patent No. 5972616
; GENERAL INFORMATION:
; APPLICANT: O'Brien, Timothy J.
; APPLICANT: Tanimoto, Hirotooshi
; TITLE OF INVENTION: TADG-15: An Extracellular Serine Protease Overexpressed in
; FILE REFERENCE: D6064
; CURRENT APPLICATION NUMBER: US/09/027,337B
; CURRENT FILING DATE: 1998-02-20
; NUMBER OF SEQ ID NOS: 13
; SEQ ID NO 1
; LENGTH: 3147
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; LOCATION: 23..2589
; OTHER INFORMATION: cDNA sequence of TADG-15
US-09-027-337-1

Alignment Scores:
Pred. No.: 1.26e-35 Length: 3147
Score: 549.00 Matches: 118
Percent Similarity: 51.80% Conservative: 40
Best Local Similarity: 38.69% Mismatches: 103
Query Match: 28.11% Indels: 44
DB: Gaps: 8

US-10-037-417-46 (1-357) x US-09-027-337-1 (1-3147)
QY 32 AlaArgGlyProProTyrCysGlyArgProGluPro----- 43
Db 1763 AGCAAGGGCAACCTGAGTGTGACGGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1822
QY 44 -----SerAlaArgIleValGlyGlySerAsn 52
Db 1823 GACTGCGACTGTGGGCTGCGTCAATTCAGCAGACAGGCTGTGTGTGGGGCAGCGAT 1882
QY 53 AlaGlnProGlyThrTrpProTrpGlnValSerLeuHisHis-----GlyGlyGlyHisIle 71
Db 1883 GCGGATGAGGGCGAGTGGCCCTGGCAGTAGCTGCTGCTGCGCCAGGCGGCACATC 1942
QY 72 CysGlyGlySerLeuIleAlaProSerTrpValLeuSerAlaAlaHisCysPheMetThr 91
Db 1943 TGGCGTGTTCCTCTCATCTCTCCCACTGGCTGTCTCTGCGCAGCAGCTGTACATCGAT 2002
QY 92 AsnGly-----ThrLeuGluProAlaAlaGluTrpSerValLeuGlyValHisSer 109
Db 2003 GACAGAGGATTCAGGTACTCAGACCCCAAGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2062
QY 110 Gln---AspGlyProLeuAspGlyAlaHisThrArgAlaValAlaAlaIleValPro 128
Db 2063 CAGAGCCAGCGCAGCGCCCTGGGGTGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2122
```

129 AlaAsnTyrSerGlnValGluLeuGlyAlaAspLeuAlaLeuLeuAlaSerPro 148
1223 CCCTTCTTCAATGACTTACCTTCGACTATGACATCGCTGCTGGAGCTGGAGAAACCG 2182
149 AlaSerLeuGlyProAlaValTyrProValCysLeuProAlaSerHisArgPheVal 168
2183 GCAGATACAGCTCATGCTGGGCGCCATCTGCTGCGGAGCTCCCATGCTTCCCT 2242
169 HisGlyThrAlaCysTyrAlaThrGlyTyrGlyAspValGlnGluAlaAspProLeuPro 188
2243 GCCGGCAAGGCCATCTGGTTCACGGGCTGGGGACACACCCAGTATGGAGGCACTGGCGCG 2302
189 LeuProTyrValLeuGlnGluValGluLeuArgLeuGlyGluAlaThrCysGlnCys 208
2303 CTG-----ATCCTGCAAAAGGCTGAGATCCGCTGATCAACCAACACCTGGAGAAC 2356
209 LeuTyrSerGlnProGlyProPheAsnLeuThrLeuGlnLeuLeuProGlyMetLeuCys 228
2357 CTCCTGCCGCGAG-----CAGATCACCGCGCGCATGATGTGC 2392
229 AlaGlyTyrProGluGlyArgArgAspThrCysGlnGlyAspSerGlyGlyProLeu--- 247
2393 GGGGCTTCTCAGCGCGCGCTGACTCTCCAGGGGTGATTCGGGGGAGCCCTGTCC 2452
248 ValCysGluGluGlyGlyArgTyrPheGlnAlaGlyIleThrSerPheGlyPheGlyCys 267
2453 AGCGTGGAGCGGATGGCGGATCTTCCAGGCGGTGTGTGAGCTGGGAGACGGCTGC 2512
268 GlyArgArgAsnArgProGlyValPheThrAlaValAlaThrTyrGluAlaTyrPheArg 287
2513 GCTCAGAGAACAGCAGCGGTGTACACAGGCTCCCTCTGTTCCGGGACTGGATCAAA 2572
288 GluGlnValMetGlySerGluProGlyProAlaPheProThrGlnProGlnLysThrGln 307
2573 GAGAACACTGGGGTATAGGGCGCGGGGCCA----- 2602
308 SerAspCysLeuHis 312
2603 CCCAAATGTGTACAC 2617

RESULT 9

US-09-644-600-1
; Sequence 1, Application US/09644600
; Patent No. 6451500
; GENERAL INFORMATION:
; APPLICANT: O'Brien, Timothy J.
; APPLICANT: Tanimoto, Hirotooshi
; TITLE OF INVENTION: TADG-15: An Extracellular Serine Protease
; TITLE OF INVENTION: Overexpressed in Carcinomas
; FILE REFERENCE: D6064CIP/D
; CURRENT APPLICATION NUMBER: US/09/644,600
; CURRENT FILING DATE: 2000-08-23
; PRIORITY APPLICATION NUMBER: 09/421,213
; PRIOR FILING DATE: 1999-10-20
; PRIOR APPLICATION NUMBER: 09/027,337
; PRIOR FILING DATE: 1998-02-20
; NUMBER OF SEQ ID NOS: 98
; SEQ ID NO 1
; LENGTH: 3147
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: TADG-15
US-09-644-600-1

Alignment Scores:
Pred. No.: 1,268-35 Length: 3147
Score: 549.00 Matches: 118
Percent Similarity: 51.80% Conservative: 40
Best Local Similarity: 38.69% Mismatches: 103
Query Match: 28.11% Indels: 44
DB: Gaps: 8

US-10-037-417-46 (1-357) x US-09-644-600-1 (1-3147)
32 AlaArgGlyProTyrCysGlyArgProGluPro----- 43
1763 AGCAAGGCAACCTTGAGTGTGACGGGAAGGAGACTGTAGCGACGGCTCAGATGAGAAG 1822
44 -----SerAlaArgIleValGlyGlySerAsn 52
1823 GACTGCGACTGTGGGCTCGGCTCATTCACGAGACAGGCTGTGTGTGTGGGGCAGCGAT 1882
53 AlaGlnProGlyThrTyrProTyrGlnValSerLeuHisHis---GlyGlyGlyHisIle 71
1883 GCGGATGAGGGCGAGTGGCCCTGCGAGGTAGCTGCTGCTGGGGCCAGGGCCACATC 1942
72 CysGlyGlySerLeuLeuAlaProSerTyrValLeuSerAlaAlaHisCysPheMetThr 91
1943 TGGGTTGTTCCCTCATCTCTCCCACTGGGTCTCTCGCCGACACTGTATACATCAT 2002
92 AsnGly-----ThrLeuGluProAlaAlaGluTyrSerValLeuLeuGlyValHisSer 109
2003 GACAGAGATTCAGGTACTCAGACCCCGAGTGGAGCGGCTTCTGGGCTTGACGAC 2062
110 Gln---AspGlyProLeuAspGlyAlaHisThrArgAlaValAlaAlaIleValPro 128
2063 CAGACCCAGCGCAGCGCCCTGGGGTGCAGGAGCGAGGCTCAAGCGCATCATCTCCAC 2122
129 AlaAsnTyrSerGlnValGluLeuGlyAlaAspLeuAlaLeuLeuArgLeuAlaSerPro 148
2123 CCCTTCTTCAATGACTTCACTTCGACTATGATCATCGGCTCTGTGAGCTGGAGAACCG 2182
149 AlaSerLeuGlyProAlaValTyrProValCysLeuProArgAlaSerHisArgPheVal 168
2183 GCAGAGTACAGCTCCATGTGGCGCCATCTGCTGCGGAGCGCTCCCATGTCTTCCCT 2242
169 HisGlyThrAlaCysTyrAlaThrGlyTyrGlyAspValGlnGluAlaAspProLeuPro 188
2243 GCCGGCAAGCCATCTGGGTACGGGCTGGGGACACACCCAGTATGGAGGCACTGGCGCG 2302
189 LeuProTyrValLeuGlnGluValGluLeuArgLeuGlyGluAlaThrCysGlnCys 208
2303 CTG-----ATCCTGCAAAAGGCTGAGATCCGCTCATCAACCAACACCTTCGAGAAC 2356
209 LeuTyrSerGlnProGlyProPheAsnLeuThrLeuGlnLeuLeuProGlyMetLeuCys 228
2357 CTCCTGCCGCGAG-----CAGATCACCGCGCGCATGATGTGC 2392
229 AlaGlyTyrProGluGlyArgArgAspThrCysGlnGlyAspSerGlyGlyProLeu--- 247
2393 GTGGGCTTCTCAGCGCGCGGTGAGTCTCTGCGAGGCTGATTCGGGGGAGCCCTGTCC 2452
248 ValCysGluGluGlyGlyArgTyrPheGlnAlaGlyIleThrSerPheGlyPheGlyCys 267
2453 AGCGTGGAGCGGATGGCGGATCTTCCAGGCGGTGTGTGAGCTGGGAGACGGCTGC 2512
268 GlyArgArgAsnArgProGlyValPheThrAlaValAlaThrTyrGluAlaTyrPheArg 287
2513 GCTCAGAGAACAGCAGCGGTGTACACAGGCTCCCTCTGTTCCGGGACTGGATCAAA 2572
288 GluGlnValMetGlySerGluProGlyProAlaPheProThrGlnProGlnLysThrGln 307
2573 GAGAACACTGGGGTATAGGGCGCGGGGCCA----- 2602
308 SerAspCysLeuHis 312
2603 CCCAAATGTGTACAC 2617

RESULT 10
US-09-644-600-18/c
; Sequence 18, Application US/09644600
; Patent No. 6451500
; GENERAL INFORMATION:
; APPLICANT: O'Brien, Timothy J.
; APPLICANT: Tanimoto, Hirotooshi

```

; TITLE OF INVENTION: TADG-15: An Extracellular Serine Protease
; FILE REFERENCE: D6064CIP/D
; CURRENT APPLICATION NUMBER: US/09/644,600
; CURRENT FILING DATE: 2000-08-23
; PRIOR APPLICATION NUMBER: 09/421,213
; PRIOR FILING DATE: 1999-10-20
; PRIOR FILING DATE: 09/027,337
; PRIOR FILING DATE: 1998-02-20
; NUMBER OF SEQ ID NOS: 98
; SEQ ID NO 18
; LENGTH: 3147
; TYPE: RNA
; ORGANISM: Artificial sequence
; FEATURE:
; OTHER INFORMATION: Antisense of TADG-15
US-09-644-600-18

Alignment Scores:
Pred. No.: 1,26e-35 Length: 3147
Score: 549.00 Matches: 118
Percent Similarity: 51.80% Conservative: 40
Best Local Similarity: 38.69% Mismatches: 103
Query Match: 28.11% Indels: 44
Gaps: 8
DB:

US-10-037-417-46 (1-357) x US-09-644-600-18 (1-3147)
QY 32 AlaArgGlyProProTyrCysGlyArgProGluPro----- 43
Db 1385 AGCAAGGGCAACCTCAGTGTGACGGAGGAGGAGCTGTAGCGACGGCTCAGATCAGAG 1326
QY 44 -----SerAlaArgIleValGlyGlySerAsn 52
Db 1325 GACTGCGACTGTGGGTGGCGTCAATTCAGACAGACGCTCGTGTGTGGGGCGACGGAT 1266
QY 53 AlaGlnProGlyThrTrpProTrpGlnValSerLeuHisHis-----GlyGlyGlyHisIle 71
Db 1265 GCGGATGAGGGCGAGTGGCCCTGGCAGGTAAGCCTGCATGCTGTGGGCCGAGGCGCCACATC 1206
QY 72 CysGlyGlySerLeuIleAlaProSerTrpValLeuSerAlaAlaHisCysPheMetThr 91
Db 1205 TGGGGTGTCTCCCTCATCTCCCAACTGGCTGTCTGCGCGCACACTGCTACATCGAT 1146
QY 92 AsnGly-----ThrLeuGluProAlaAlaGluTrpSerValLeuLeuGlyValHisSer 109
Db 1145 GACAGAGGATTCAGGTACTCAGACCCACCCAGTGGAGCGCCTCTCGGCTTGCAGCAG 1086
QY 110 Gln---AspGlyProLeuAspGlyAlaHisThrArgAlaValAlaIleValValPro 128
Db 1085 CAGAGCCAGCGCAGCGCCCTGGGGTGCAGGAGCGCAGGCTCAAGCGCATCATCTCCAC 1026
QY 129 AlaAsnTyrSerGlnValGluLeuGlyAlaAspLeuAlaLeuArgLeuAlaSerPro 148
Db 1025 CCCTTCTCAATGACTTCACCTTCGACTATGACATCGCGCTGTCTGGAGCTGGAGAACCG 966
QY 149 AlaSerLeuGlyProAlaValTrpProValCysLeuProArgAlaSerHisArgPheVal 168
Db 965 GCAGGTACAGCTCATCGTGGTGGGCCCATCTGCTGCGGAGCGCTCCCATGTCTCCCT 906
QY 169 HisGlyThrAlaCysTrpAlaThrGlyTrpGlyAspValGlnGluAlaAspProLeuPro 188
Db 905 GCCGCAAGGCCCATCTGGGTACCGGCTGGGGACACACCCAGTATGGAGGCACCTGGCGCG 846
QY 189 LeuProTrpValLeuGlnGluValGluLeuArgLeuLeuGlyGluAlaThrCysGlnCys 208
Db 845 CTG-----ATCCTGCAAAAGGTGAGATCCGGTCAACACCCAGACCCATCTGGAGAAC 792
QY 209 LeuTyrSerGlnProGlyProPheAsnLeuThrLeuGlnIleLeuProGlyMetLeuCys 228
Db 791 CTCCTGCCGCGAG-----CAGATCACCGCCCGCATGATGTCG 756
QY 229 AlaGlyTyrProGluGlyArgAspThrCysGlnGlyAspSerGlyGlyProLeu---- 247
```

```

Db 755 GTGGGCTTCTCAGCGCGCGCGTGGACTCTCCAGGGTGATTCGGGGGACCCCTGTCC 696
QY 248 ValCysGluGluGlyGlyArgTrpPheGlnAlaGlyIleThrSerPheGlyPheGlyCys 267
Db 695 AGCGTGGAGCGGATGGCGGATCTCCAGGCCGGTGTGGTGGAGTGGGAGACGGCTGC 636
QY 268 GlyArgArgAsnArgProGlyValPheThrAlaValAlaThrTyrGluAlaTrpIleArg 287
Db 635 GCTCAGAGGAACAAGCCAGCGGTGTACACAAGGCTCCCTCTGTTTCGGGACTGGATCAA 576
QY 288 GluGlnValMetGlySerGluProGlyProAlaPheProThrGlnProGlnLysThrGln 307
Db 575 GAGAACACTGGGGTATAGGGGCGCGGCCCA----- 546
QY 308 SerAspCysLeuHis 312
Db 545 CCCAATGTGTACAC 531

RESULT 11
US-09-654-600A-1
; Sequence 1, Application US/09654600A
; Patent No. 6649741
; GENERAL INFORMATION:
; APPLICANT: O'Brien, Timothy J.
; APPLICANT: Tanimoto, Hirotooshi
; TITLE OF INVENTION: TADG-15: An Extracellular Serine Protease
; FILE REFERENCE: D6064CIP/D
; CURRENT APPLICATION NUMBER: US/09/654,600A
; CURRENT FILING DATE: 2000-09-01
; PRIOR APPLICATION NUMBER: 09/421,213
; PRIOR FILING DATE: 1999-10-20
; NUMBER OF SEQ ID NOS: 98
; SEQ ID NO 1
; LENGTH: 3147
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: TADG-15
US-09-654-600A-1

Alignment Scores:
Pred. No.: 1,26e-35 Length: 3147
Score: 549.00 Matches: 118
Percent Similarity: 51.80% Conservative: 40
Best Local Similarity: 38.69% Mismatches: 103
Query Match: 28.11% Indels: 44
Gaps: 8
DB:

US-10-037-417-46 (1-357) x US-09-654-600A-1 (1-3147)
QY 32 AlaArgGlyProProTyrCysGlyArgProGluPro----- 43
Db 1763 AGCAAGGGCAACCTCAGTGTGACGGAGGAGGAGCTGTAGCGACGGCTCAGATGAGAAG 1822
QY 44 -----SerAlaArgIleValGlyGlySerAsn 52
Db 1823 GACTGCGACTGTGGGTGGCGTCAATTCAGACAGACGCTGTGTGTGGGGCAGCGAT 1882
QY 53 AlaGlnProGlyThrTrpProTrpGlnValSerLeuHisHis-----GlyGlyGlyHisIle 71
Db 1883 GCGGATGAGGGCGGAGTGGCCCTGGCAGGTAGAGCTGTGATCTCTGGGCCAGGCGCCATC 1942
QY 72 CysGlyGlySerLeuIleAlaProSerTrpValLeuSerAlaAlaHisCysPheMetThr 91
Db 1943 TGGGGTGTCTCCCTCATCTCTCCCAACTGGCTGTCTCTCGCGCACACTGCTACATCGAT 2002
QY 92 AsnGly-----ThrLeuGluProAlaAlaGluTrpSerValLeuLeuGlyValHisSer 109
Db 2003 GACAGAGGATTCAGGTACTCAGACCCACCGAGTGGACGCTTCTTGGGGCTTGCACGAC 2062
```


QY 110 Gln---AspGlyProLeuAspGlyAlaHisThrArgAlaValAlaAlaValValValPro 128
 Db 2063 CAGAGCCAGGCGAGCCCTGGGGTGCAGGAGCGAGGCTCAAGCGCATCATCTCCAC 2122
 QY 129 AlaAsnTyrSerGlnValGluLeuGlyAlaAspLeuAlaLeuArgLeuAlaSerPro 148
 Db 2123 CCCTTCTCAATGACTTCACCTTCGACTATGACATCGCGCTGCTGAGCTGGAGAACCG 2182
 QY 149 AlaSerLeuGlyProAlaValTrpProValCysLeuProArgAlaSerHisArgPheVal 168
 Db 2183 GCAGAGTACAGCTCCATGGTGGCGCCATCTGCTCGCGAGCCCTCCCATGCTTCCCT 2242
 QY 169 HisGlyThrAlaCysTrpAlaThrGlyTrpGlyAspValGlnGluAlaAspProLeuPro 188
 Db 2243 GCCGCAAGGCCATCTGGGTCAGCGGCTGGGGACACACCAGTATGGAGGCATGGCGCG 2302
 QY 189 LeuProTrpValLeuGlnGluValGluLeuArgLeuLeuGlyGluAlaThrCysGlnCys 208
 Db 2303 CTG-----ATCCTGCAAAAGGTGAGATCGGCTCATCAACAGACCACCTGGAGAAC 2356
 QY 209 LeuTyrSerGlnProGlyProPheAsnLeuThrLeuGlnLeuProGlyMetLeuCys 228
 Db 2357 CTCTGCGCGAG-----CAGATCAGCGCGCATGATGTC 2392
 QY 229 AlaGlyTyrProGluGlyArgArgAspThrCysGlnGlyAspSerGlyGlyProLeu--- 247
 Db 2393 GTGGGCTTCTCAGCGCGCGTGGTACTCTCGCAGGGTGAATCCGGGGGACCCCTGTC 2452
 QY 248 ValCysGluGluGlyGlyArgTrpPheGlnAlaGlyIleThrSerPheGlyPheGlyCys 267
 Db 2453 AGCTGAGGCGGATGGCGGATCTTCCAGCGCGTGTGGTGTGAGCTGGGAGAGCGGTGC 2512
 QY 268 GlyArgAsnArgProGlyValPheThrAlaValAlaThrTyrGluAlaTrpIleArg 287
 Db 2513 GCTCAGAGGAACAAGCAGCGGCTGTACACAAGGCTCCTCTGTTTCGGGACTGGATCAAA 2572
 QY 288 GluGlnValMetGlySerGluProGlyProAlaPheProThrGlnProGlnLysThrGln 307
 Db 2573 GAGAACACTGGGGTATAGGGCGCGGGCCA----- 2602
 QY 308 SerAspCysLeuHis 312
 Db 2603 CCCAAATGTGTACAC 2617

RESULT 12

; Sequence 18, Application US/09654600A
 ; Patent No. 6649741
 ; GENERAL INFORMATION:
 ; APPLICANT: O'Brien, Timothy J.
 ; APPLICANT: Tanimoto, Hirotochi
 ; TITLE OF INVENTION: TADG-15: An Extracellular Serine Protease
 ; TITLE OF INVENTION: Overexpressed in Carcinomas
 ; FILE REFERENCE: D6064CIP/D
 ; CURRENT APPLICATION NUMBER: US/09/654,600A
 ; CURRENT FILING DATE: 2000-09-01
 ; PRIOR APPLICATION NUMBER: 09/421,213
 ; 09/027,337
 ; PRIOR FILING DATE: 1999-10-20
 ; 1998-02-20
 ; NUMBER OF SEQ ID NOS: 98
 ; SEQ ID NO 18
 ; LENGTH: 3147
 ; TYPE: RNA
 ; ORGANISM: Artificial sequence
 ; FEATURE:
 ; OTHER INFORMATION: Antisense of TADG-15
 US-09-654-600A-18

Alignment Scores: 1.26e-35 Length: 3147
 Pred. No.: 549.00 Matches: 118
 Score:

Percent Similarity: 51.80% Conservative: 40
 Best Local Similarity: 38.69% Mismatches: 103
 Query Match: 28.11% Indels: 44
 Db: 4 Gaps: 8
 US-10-037-417-46 (1-357) x US-09-654-600A-18 (1-3147)
 QY 32 AlaArgGlyProProTyrCysGlyArgProGluPro----- 43
 Db 1385 AGCAAGGCAACCTGAGTGTGACGGAAGGAGGACTGTAGCGAGCGTTCAGATGAGAAG 1326
 QY 44 -----SerAlaArgIleValGlyGlySerAsn 52
 Db 1325 GACTGCGACTGGGCTCGGTCATTACAGACAGGCTCGTGTGTTGGGGCAGCGAT 1266
 QY 53 AlaGlnProGlyThrTrpProTrpGlnValSerLeuHisHis---GlyGlyGlyHisIle 71
 Db 1265 GCGATGAGGGGAGTGGCCCTGGCAGTAAAGCTGCTGCTGCGCAGGCGCACATC 1206
 QY 72 CysGlyGlySerLeuIleAlaProSerTrpValLeuSerAlaAlaHisCysPheMetThr 91
 Db 1205 TSCGGTGTCTCCCTCATCTCTCCAACTGGTGTCTCTGCGCACACATCTCATCGAT 1146
 QY 92 AsnGly-----ThrLeuGluProAlaAlaGluTrpSerValLeuLeuGlyValHisSer 109
 Db 1145 GACAGAGGATTCAGGTACTCAGACCCACGCGAGTGGAGCGSCCTTCTGGGCTTGCAGAC 1086
 QY 110 Gln---AspGlyProLeuAspGlyAlaHisThrArgAlaValAlaAlaIleValValPro 128
 Db 1085 CAGAGCCAGGCGAGCCCTGGGTCAGGAGCGGAGCTCAAGCGCATCATCTCCAC 1026
 QY 129 AlaAsnTyrSerGlnValGluLeuGlyAlaAspLeuAlaLeuLeuArgLeuAlaSerPro 148
 Db 1025 CCCTTCTCAATGACTTCACCTTCGACTATGATGATGCTGCTGGAGCTGGAGAACCG 966
 QY 149 AlaSerLeuGlyProAlaValTrpProValCysLeuProArgAlaSerHisArgPheVal 168
 Db 965 GCAGAGTACAGCTCCATGGTGGCGCCATCTGCTGCGGAGCGCTCCATCTTCCCT 906
 QY 169 HisGlyThrAlaCysTrpAlaThrGlyTrpGlyAspValGlnGluAlaAspProLeuPro 188
 Db 905 GCCGCAAGGCCATCTGGTTCAGCGGCTGGGGACACACCAGTATGGAGGCATCTGGCGCG 846
 QY 189 LeuProTrpValLeuGlnGluValGluLeuArgLeuLeuGlyGluAlaThrCysGlnCys 208
 Db 845 CTG-----ATCCTGCAAAAGGTGAGATCGCGCTCATCAACAGACACCTGGAGAAC 792
 QY 209 LeuTyrSerGlnProGlyProPheAsnLeuThrLeuGlnIleLeuProGlyMetLeuCys 228
 Db 791 CTCTGCGCGAG-----CAGATCAGCGCGCATGATGTGC 756
 QY 229 AlaGlyTyrProGluGlyArgArgAspThrCysGlnGlyAspSerGlyGlyProLeu--- 247
 Db 755 GTGGGCTTCTCAGCGCGCGTGGTACTCTGCGCAGGGTGAATCCGGGGGACCCCTGTCC 696
 QY 248 ValCysGluGluGlyGlyArgTrpPheGlnAlaGlyIleThrSerPheGlyPheGlyCys 267
 Db 695 AGCTGAGGCGGATGGGCGGATCTTCCAGCGCGGTGTGGTGTGAGTGGGAGACCGGCTGC 636
 QY 268 GlyArgArgAsnArgProGlyValPheThrAlaValAlaThrTyrGluAlaTrpIleArg 287
 Db 635 GCTCAGAGGAACAAGCAGCGGCTGTACACAAGGCTCCCTCTGTTTCGGGACTGGATCAAA 576
 QY 288 GluGlnValMetGlySerGluProGlyProAlaPheProThrGlnProGlnLysThrGln 307
 Db 575 GAGAACACTGGGGTATAGGGCGCGGGCCA----- 546
 QY 308 SerAspCysLeuHis 312
 Db 545 CCCAAATGTGTACAC 531

RESULT 13
 US-09-907-794A-262

Db 955 TGGGGCCCTCAGGACCGACCGAGGCT-CTGGGGCGCGCCGCGCTCTAGGCGCGAG 1013
 QY 308 SerAspCysLeuHisGlnThrAlaPheLeuAspSerAlaArgIleLeuLeuArgProLeu 327
 Db 1014 CGGAC-----CGGGGCTCGGATCTGAAGG-----CGGCGAGAT 1049
 QY 328 SerHisIleSerVal-----GlyValSerThrGlyThrIlySerLeu 341
 Db 1050 CCACATCTGGATCTGGATCTGGCGGCTCTGGCGGTTTCCCGCGCGTAATAAGGCTC 1109
 QY 342 ValLeuProTrpLeuSerProHisSerLeuLeuGlyLeuTrpGly 356
 Db 1110 ATC-----TACCTTACCTCTGGGG 1130

RESULT 14

US-09-905-125A-262
 ; Sequence 262, Application US/09905125A
 ; Patent No. 6664376
 ; GENERAL INFORMATION:
 ; APPLICANT: Genentech, Inc.
 ; APPLICANT: Ashkenazi, Avi
 ; APPLICANT: Botstein, David
 ; APPLICANT: Desnoyers, Luc
 ; APPLICANT: Eaton, Dan L.
 ; APPLICANT: Ferrara, Napoleone
 ; APPLICANT: Filvaroff, Ellen
 ; APPLICANT: Fong, Sherman
 ; APPLICANT: Gao, Wei-Qiang
 ; APPLICANT: Gerber, Hanspeter
 ; APPLICANT: Gerritsen, Mary E.
 ; APPLICANT: Goddard, A.
 ; APPLICANT: Godowski, Paul J.
 ; APPLICANT: Grimaldi, Christopher J.
 ; APPLICANT: Gurney, Austin L.
 ; APPLICANT: Hillan, Kenneth, J.
 ; APPLICANT: KJavin, Ivar J.
 ; APPLICANT: Kather, Jennie P.
 ; APPLICANT: Pan, James
 ; APPLICANT: Paoni, Nicholas F.
 ; APPLICANT: Roy, Margaret Ann
 ; APPLICANT: Stewart, Timothy A.
 ; APPLICANT: Tumas, Daniel
 ; APPLICANT: Williams, P. Mickey
 ; APPLICANT: Wood, William, I.
 ; TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
 ; FILE REFERENCE: 10466-14
 ; CURRENT APPLICATION NUMBER: US/09/905,125A
 ; PRIOR FILING DATE: 2001-07-12
 ; PRIOR APPLICATION NUMBER: PCT/US00/04414
 ; PRIOR FILING DATE: 2000-02-22
 ; PRIOR APPLICATION NUMBER: US 60/143,048
 ; PRIOR FILING DATE: 1999-07-07
 ; PRIOR APPLICATION NUMBER: US 60/145,698
 ; PRIOR FILING DATE: 1999-07-26
 ; PRIOR APPLICATION NUMBER: US 60/146,222
 ; PRIOR FILING DATE: 1998-07-28
 ; PRIOR APPLICATION NUMBER: PCT/US99/20594
 ; PRIOR FILING DATE: 1999-09-08
 ; PRIOR APPLICATION NUMBER: PCT/US99/20944
 ; PRIOR FILING DATE: 1999-09-13
 ; PRIOR APPLICATION NUMBER: PCT/US99/21090
 ; PRIOR FILING DATE: 1999-09-15
 ; PRIOR APPLICATION NUMBER: PCT/US99/21547
 ; PRIOR FILING DATE: 1999-09-15
 ; PRIOR APPLICATION NUMBER: PCT/US99/23089
 ; PRIOR FILING DATE: 1999-10-05
 ; PRIOR APPLICATION NUMBER: PCT/US99/28214
 ; PRIOR FILING DATE: 1999-11-29
 ; PRIOR APPLICATION NUMBER: PCT/US99/28313
 ; PRIOR FILING DATE: 1999-11-30
 ; PRIOR APPLICATION NUMBER: PCT/US99/28564
 ; PRIOR FILING DATE: 1999-12-02

; PRIOR APPLICATION NUMBER: PCT/US99/28565
 ; PRIOR FILING DATE: 1999-12-02
 ; PRIOR APPLICATION NUMBER: PCT/US99/30095
 ; PRIOR FILING DATE: 1999-12-16
 ; PRIOR APPLICATION NUMBER: PCT/US99/30911
 ; PRIOR FILING DATE: 1999-12-20
 ; PRIOR APPLICATION NUMBER: PCT/US99/30999
 ; PRIOR FILING DATE: 1999-12-20
 ; PRIOR APPLICATION NUMBER: PCT/US00/00219
 ; PRIOR FILING DATE: 2000-01-05
 ; NUMBER OF SEQ ID NOS: 423
 ; SEQ ID NO 262
 ; LENGTH: 1378
 ; TYPE: DNA
 ; ORGANISM: Homo Sapien
 ; US-09-905-125A-262
 Alignment Scores:
 Pred. No.: 8,75e-36 Length: 1378
 Score: 545.00 Matches: 134
 Percent Similarity: 48.00% Conservatives: 46
 Best Local Similarity: 35.73% Mismatches: 146
 Query Match: 27.91% Indels: 51
 DB: 4 Gaps: 9
 US-10-037-417-46 (1-357) x US-09-905-125A-262 (1-1378)
 QY 7 LeuGlyProGlyGlnLeuGlyAlaValAlaAsnSerAspSerTyrSerLeuTyrGlyLeu 26
 Db 80 CTGGGTGGGGCTGTCTCGGCACCTTCACCTCCCTCTGCTGCTGGCGTGCAGAGCCATC 139
 QY 27 ValProSerGlyProAlaArgGlyProProTyrCysGlyArgProGluProSerAlaArg 46
 Db 140 CTCATGCGCGCCAGGATACCTGTTCCTCCCGACCTGTGGGAAGCCCGCAGCGCTGAACCG 199
 QY 47 IleValGlyGlySerAsnAlaGlnProGlyThrTrpProTrpGlnValSerLeuHisHis 66
 Db 200 GTTGTGGCGCGGAGGAGGACGACTGACGCGAGTGGCCCTGGATCGTGAGCATCCAGAAG 259
 QY 67 GlyGlyGlyHisIleCysGlyGlySerLeuIleAlaProSerTrpValLeuSerAlaAla 86
 Db 260 AATGGGACCCACCACTCGCGAGGTTCTCTGCTCACCAGCCGCTGGGTGATCCTCTGCTGCC 319
 QY 87 HisCysPheMetThrAsnGlyThrLeuGluProAlaAlaGluTrpSerValLeuLeuGly 106
 Db 320 CACTGTTCGAAGGACAC-----CTGAACAACCATACCTGTTCCTCTGCTGCTGCTGCC 373
 QY 107 ValHisSerGlnAspGlyProLeuAspGlyAlaHisThrArgAlaValAlaAlaIleVal 126
 Db 374 GCCTGGCAGCTGGGGAACCTGGCTCTCGGTCCAGAGGTGGGTGTTGCTGGGTGGAG 433
 QY 127 ValProAlaAsnTyrSerGlnValGlu-----LeuGlyAlaAspLeuAlaLeuLeuArgLeu 145
 Db 434 CCCACCTCTGTGTATTCTCTGAAGGAAGTCTGTGTGCAGACATTCCTGCTGGTGGTCTC 493
 QY 146 AlaSerProAlaSerLeuGlyProAlaValTrpProValCysLeuProArgAlaSerHis 165
 Db 494 GAGCGCTCCATACAGTTCTCTCAGAGCGGTCTCGCCATCTGCCCTACCTACCTGCTATC 553
 QY 166 ArgPheValHisGlyThrAlaCysTrpAlaThrGlyTrpGlyAspValGlnGlnAlaAsp 185
 Db 554 CACTCCCTCCAAACACCACTGCTGGATCTCAGGCTGGGAGCATCAAGATGGAGTT 613
 QY 186 ProLeuProLeuProTrpValLeuGlnValGluLeuArgLeuLeuGlyGluAlaThr 205
 Db 614 CCTTGGCCCTCCAGACCTCGAGAGGTTCCTATCATCATCATCATCATCATCATCATCAT 673
 QY 206 CysGlnCysLeuTyr-----SerGlnProGlyProPheAsnLeuThrLeuGlnIle 222
 Db 674 TGCAGCCCATCTGTACTGGCGGGGAGCAGGAGGACCC-----ATC 715
 QY 223 LeuProGlyMetLeuCysAlaGlyTyrProGluGlyArgArgAspThrCysGlnGlyAsp 242

Thu Feb 26 13:50:27 2004

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - nucleic search, using frame_plus.p2n model

Run on: February 25, 2004, 17:20:05 ; Search time 399 Seconds
(without alignments)
3133.050 Million cell updates/sec

Title: US-10-037-417-46

Perfect score: 1953

Sequence: 1 MAQKGVLPQGLGAVANSDS.....TKSLVLPWSPHSLGLNGF 357

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 2308694 seqs, 1750822206 residues

Total number of hits satisfying chosen parameters: 4617368

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame.p2n.model -DEV=xlp
-Q=/cgn2_1/USPTO.spool_p/US10037417/runat_24022004_101040_17563/app_query.fasta_1.519
-DB=Published Applications NA -QFMT=fastap -SUFFIX=p2n.rnpb -MINMATCH=0.1
-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62
-TRANS-human40.cdi -LIST=45 -DOCALIGN=200 -THR SCORE=pct -THR MAX=100
-THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0
-MAXLEN=2000000000 -USER=US10037417@cgn 1 107 @runat 24022004_101040_17563
-NCPU=6 -ICPU=3 -NO_MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100
-LONGLOG -DEV TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5
-FGAPOP=6 -FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published Applications NA:
1: /cgn2_6/ptodata/1/pubna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/1/pubna/PCT_NEW_PUB.seq:*
3: /cgn2_6/ptodata/1/pubna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/1/pubna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/1/pubna/PCTUS_PUBCOMB.seq:*
6: /cgn2_6/ptodata/1/pubna/US08_NEW_PUB.seq:*
7: /cgn2_6/ptodata/1/pubna/US08_PUBCOMB.seq:*
8: /cgn2_6/ptodata/1/pubna/US09A_PUBCOMB.seq:*
9: /cgn2_6/ptodata/1/pubna/US09B_PUBCOMB.seq:*
10: /cgn2_6/ptodata/1/pubna/US09C_PUBCOMB.seq:*
11: /cgn2_6/ptodata/1/pubna/US09_NEW_PUB.seq:*
12: /cgn2_6/ptodata/1/pubna/US10A_PUBCOMB.seq:*
13: /cgn2_6/ptodata/1/pubna/US10B_PUBCOMB.seq:*
14: /cgn2_6/ptodata/1/pubna/US10C_PUBCOMB.seq:*
15: /cgn2_6/ptodata/1/pubna/US10_NEW_PUB.seq:*
16: /cgn2_6/ptodata/1/pubna/US06_NEW_PUB.seq:*
17: /cgn2_6/ptodata/1/pubna/US06_PUBCOMB.seq:*
18: /cgn2_6/ptodata/1/pubna/US06_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description

1	1505	77.1	2457	9	US-09-888-615-52	Sequence 52, Appl
2	697	35.7	1733	14	US-10-176-847-85	Sequence 85, Appl
3	697	35.7	1834	9	US-09-948-094-1	Sequence 1, Appl
4	697	35.7	1834	9	US-09-880-107-2214	Sequence 2214, Ap
5	697	35.7	1834	9	US-09-967-768A-141	Sequence 141, App
6	697	35.7	1834	14	US-10-097-340-261	Sequence 261, App
7	697	35.7	3382	14	US-10-101-510-447	Sequence 447, App
8	665	34.1	1668	9	US-09-925-301-208	Sequence 208, App
9	652	33.4	1726	12	US-10-042-865-29	Sequence 29, Appl
10	627.5	32.1	1797	14	US-10-109-616-1	Sequence 1, Appl
11	612	31.3	1161	12	US-10-042-865-31	Sequence 31, Appl
12	603.5	30.9	1613	13	US-10-041-400A-1	Sequence 1, Appl
13	603.5	30.9	1613	13	US-10-041-264A-1	Sequence 1, Appl
14	603.5	30.9	1613	13	US-10-042-091A-1	Sequence 25, Appl
15	591	30.3	1020	15	US-10-051-874-25	Sequence 29, Appl
16	582.5	29.8	1958	16	US-10-311-035-29	Sequence 10, Appl
17	575.5	29.5	768	14	US-10-221-097-10	Sequence 33, Appl
18	575.5	29.5	882	12	US-10-042-865-33	Sequence 34, Appl
19	571.5	29.3	882	12	US-10-042-865-34	Sequence 27, Appl
20	565	28.9	1887	9	US-09-888-615-27	Sequence 15, Appl
21	565	28.9	1973	14	US-10-190-030B-15	Sequence 1, Appl
22	563.5	28.9	1110	13	US-10-040-655-1	Sequence 1, Appl
23	563.5	28.9	1110	13	US-10-041-006A-1	Sequence 221, App
24	563.5	28.9	1129	12	US-10-147-493-221	Sequence 221, App
25	563.5	28.9	1129	14	US-10-028-072-221	Sequence 221, App
26	563.5	28.9	1129	14	US-10-121-049-221	Sequence 221, App
27	563.5	28.9	1129	14	US-10-123-904-221	Sequence 221, App
28	563.5	28.9	1129	14	US-10-140-470-221	Sequence 221, App
29	563.5	28.9	1129	14	US-10-175-746-221	Sequence 221, App
30	563.5	28.9	1129	14	US-10-176-918-221	Sequence 221, App
31	563.5	28.9	1129	14	US-10-137-865-221	Sequence 221, App
32	563.5	28.9	1129	14	US-10-176-921-221	Sequence 221, App
33	563.5	28.9	1129	14	US-10-140-474-221	Sequence 221, App
34	563.5	28.9	1129	14	US-10-142-431-221	Sequence 221, App
35	563.5	28.9	1129	14	US-10-143-114-221	Sequence 221, App
36	563.5	28.9	1129	14	US-10-140-002-221	Sequence 221, App
37	563.5	28.9	1129	14	US-10-142-419-221	Sequence 221, App
38	563.5	28.9	1129	14	US-10-123-262-221	Sequence 221, App
39	563.5	28.9	1129	14	US-10-142-423-221	Sequence 221, App
40	563.5	28.9	1129	14	US-10-121-050-221	Sequence 221, App
41	563.5	28.9	1129	14	US-10-141-755-221	Sequence 221, App
42	563.5	28.9	1129	14	US-10-143-032-221	Sequence 221, App
43	563.5	28.9	1129	14	US-10-123-108-221	Sequence 221, App
44	563.5	28.9	1129	14	US-10-123-236-221	Sequence 221, App
45	563.5	28.9	1129	14	US-10-123-261-221	Sequence 221, App

ALIGNMENTS

RESULT 1
US-09-888-615-52
; Sequence 52, Application US/09888615
; Patent No. US20020064856A1
; GENERAL INFORMATION:
; APPLICANT: PLOWMAN, GREGORY
; APPLICANT: WHITE, DAVID
; APPLICANT: CAENEPEEL, SEAN
; APPLICANT: CHARYDCZAK, GLEN
; APPLICANT: MANNING, GERARD
; APPLICANT: SUPARGANAM, SUCHA
; TITLE OF INVENTION: NOVEL PROTEASES
; FILE REFERENCE: 038602/1214
; CURRENT APPLICATION NUMBER: 2001-06-26
; PRIOR FILING DATE: 2000-06-26
; PRIOR FILING DATE: 2000-06-26
; NUMBER OF SEQ ID NOS: 150
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 52
; LENGTH: 2457
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-888-615-52

; TITLE OF INVENTION: COMPOSITIONS, KITS, AND METHODS FOR
 ; TITLE OF INVENTION: IDENTIFICATION, ASSESSMENT, PREVENTION, AND THERAPY OF BREAST
 ; FILE REFERENCE: MEI-039
 ; CURRENT APPLICATION NUMBER: US/10/176,847
 ; NUMBER OF SEQ ID NOS: 112
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 85
 ; LENGTH: 1733
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 ; US-10-176-847-85

Alignment Scores:
 Pred. No.: 1,07e-126 Length: 2457
 Score: 1505.00 Matches: 272
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 77.06% Indels: 0
 DB: 9 Gaps: 0

US-10-037-417-46 (1-357) x US-09-888-615-52 (1-2457)

QY 38 CysGlyArgProGluProSerAlaArgIleValGlyGlySerAsnAlaGlnProGlyThr 57
 DB 112 TCGGGGCGCCCTGAGCCCTCGGCCGCGATCGTGGGGGCTCAACGCGAGCGGGCACC 171
 QY 58 TrpProTyrGlnValSerLeuHisGlyGlyGlyHisIleCysGlyGlySerLeuIle 77
 DB 172 TGGCCTTGGCAAGTGAAGCTGACCATGGAGTGGCCACATCTGGGGGGCTCCCTTCATC 231
 QY 78 AlaProSerTrpValLeuSerAlaAlaHisCysPheMetThrAsnGlyThrLeuGluPro 97
 DB 232 GCGCCCTCTCGGCTCTCTCCGCTGCTCACTGTTCATGACGATGGGAGCGCTGGAGCCC 291
 QY 98 AlaAlaGluTrpSerValLeuLeuGlyValHisSerGlnAspGlyProLeuAspGlyAla 117
 DB 292 GCGGCGAGTGGTGGTACTGTGGGCGTGCATCTCCAGAGCGGCGCCCTGGAGCGCGG 351
 QY 118 HisThrArgAlaValAlaAlaIleValProAlaAsnTyrSerGlnValGluLeuGly 137
 DB 352 CACACCGCGAGTGGCGCCCATCGTGGCGCGTGCATCTCCAGAGCGGCGCCCTGGAGCGG 411
 QY 138 AlaAspLeuAlaLeuLeuArgLeuAlaSerProAlaSerLeuGlyProAlaValTrpPro 157
 DB 412 GCGGACCTGGCCCTGCTGGCCCTGGCTCACCGCCGAGCTGGGCCCGCGCTGGGCT 471
 QY 158 ValCysLeuProArgAlaSerHisArgPheValHisGlyThrAlaCysTrpAlaThrGly 177
 DB 472 GTCCTGCTCGCCCGCGCTCACACCGCTTCGTGACCGGACCGCTGCTGGGCGCACCGGC 531
 QY 178 TrpGlyAspValGlnGluAlaAspProLeuProLeuProTrpValLeuGlnGluValGlu 197
 DB 532 TGGGAGAGCTCCAGAGGAGAGATCTCTGCTCTCCCTGCTGCTGCTGCTGCTGCTGCTG 591
 QY 198 LeuArgLeuLeuGlyGluAlaThrCysGlnCysLeuTyrSerGlnProGlyPropheAsn 217
 DB 592 CTAAGGCTGCTGGCGAGGCGCCACTGTCAATGCTCTACAGCGGCGCGCTCCCTTCAAC 651
 QY 218 LeuThrLeuGlnIleLeuProGlyMetLeuCysAlaGlyTyrProGluGlyArgArgAsp 237
 DB 652 CTCACCTCCAGATATTGCCAGGAGATGCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCT 711
 QY 238 ThrCysGlnGlyAspSerGlyGlyProLeuValCysGluGluGlyGlyArgTrpPheGln 257
 DB 712 ACCTGCGAGGCTGACTCTGGGGGCGCCCTGCTGTGAGGAGGAGCGGCGCTGGTCCAG 771
 QY 258 AlaGlyIleThrSerPheGlyPheGlyCysGlyArgArgAsnArgProGlyValPheThr 277
 DB 772 GCAGGATACCACTTTGGCTTTGGCTGTGGAGCGGAGAAACCGCCCTGGAGTTTCACT 831
 QY 278 AlaValAlaThrTrpGluAlaTrpIleArgGluGlnValMetGlySerGluProGlyPro 297
 DB 832 GCTGTGGCTACCTATGAGGATGATACGGGAGAGGATGATGGTTCATGAGCGCTGGGCGCT 891
 QY 298 AlaPheProThrGlnProGlnLysThrGlnSerAsp 309
 DB 892 GCCTTTCCACCCAGCCCGCCAGAGACCCAGTCAGAT 927

RESULT 2

US-10-176-847-85
 ; Sequence 85, Application US/10176847
 ; Publication No. US20030068636A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Veibay, Petter Ole

Alignment Scores:
 Pred. No.: 8.57e-54 Length: 1733
 Score: 697.00 Matches: 160
 Percent Similarity: 55.29% Conservative: 49
 Best Local Similarity: 42.33% Mismatches: 133
 Query Match: 35.69% Indels: 36
 DB: 14 Gaps: 11

US-10-037-417-46 (1-357) x US-10-176-847-85 (1-1733)

QY 1 MetAlaGlnLysGlyValLeuGlyProGlyGlnLeuGlyAlaValAlaAsnSerAspSer 20
 DB 99 ATGGCCCAAGAGGGGCTCTGGGGCTGGGAGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 146
 QY 21 TyrSerLeuTyr---GlyLeuValProSerGlyPro-----AlaArgGlyPro 35
 DB 147 ATTCTGCTCTATCTTGGATTACTCCGTCAGGACAGGAGCGGAGGAGGAGGAGGAGGAGGAGG 206
 QY 36 ProTyrCysGlyArgProGluProSerAlaArgIleValGlyGlySerAsnAlaGlnPro 55
 DB 207 -----TGCGGT---GTGGCGCCCGCCAGACGCGCATCACAGGTGGCGAGTCAGTCGCGC 257
 QY 56 GlyThrTrpProTrpGlnValSerLeuHisGlyGlyGlyHisIleCysGlyGlySer 75
 DB 258 GGTCTAGTGGCCCTGGCAGGTGAGCATCACCTATGAGGCGTCCATGTGTGTGTGTGTGTGTGT 317
 QY 76 LeuIleAlaProSerTrpValLeuSerAlaAlaHisCysPheMetThrAsnGlyThrLeu 95
 DB 318 CTCGCTCTGAGCAGCTGGGTGCTGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 377
 QY 96 GluProAlaAlaGluTrpSerValLeuLeuGlyValHisSerGlnAspGlyProLeuAsp 115
 DB 378 GAA-----GCCTATGAGGTCAAGCTGGGGGCGCCACACGCTAGACTCTCTCTCCGAG 428
 QY 116 GlyAlaHisThrArgAlaValAlaAlaIleValValProAlaAsnTyrSerGlnValGlu 135
 DB 429 GACGCCAAGGTGAGCAGCCCTGAGGAGCATCATCCCGCCCGCCCGCCCGCCCGCCCGCCCG 488
 QY 136 LeuGlyAlaAspLeuAlaLeuArgLeuAlaSerProAlaSerLeuGlyProAlaVal 155
 DB 489 TCCCGAGCGCGACATTGCACTCTCTCCAACTCAGCAGACCCATCACCTTCTCCGCTATCATC 548
 QY 156 TrpProValCysLeuProArgAlaSerHisArgPheValHisGlyThrAlaCysTrpAla 175
 DB 549 CGGCCCATCTGCTCTCCGAGCCAAACGCTCTCTCCCAACGCGCTCTCTCCGCTATCATC 608
 QY 176 ThrGlyTrpGlyAspValGlnGluAlaAspProLeuProLeuProTrpValLeuGlnGlu 195
 DB 609 ACTGGCTGGGGTCTATGTGGCCCTCTGAGTGGAGCTCTCTGAGCGCCCGCCCGCCCGCCCG 668
 QY 196 ValGluLeuArgLeuLeuGlyGluAlaThrCysGlnCysLeuTyrSerGlnProGlyPro 215
 DB 669 CTCGAGTGGCTCTGATCATCTGATCATCTGATCATCTGATCATCTGATCATCTGATCATCTG 728
 QY 216 PheAsnLeuThrLeuGlnIleLeuProGlyMetLeuCysAlaGlyTyrProGluGlyArg 235
 DB 729 CCTGAGGAGCGGACCTTTGTCCAGAGGAGCATGTGTGTGTGTGTGTGTGTGTGTGTGTGT 788
 QY 236 ArgAspThrCysGlnGlyAspSerGlyGlyProLeuValCysGluGluGlyGlyArgTrp 255

```

Db      789  AAGGAGCGCTGCAGGTCAGCTCTGGGGGCCACCTCTCCCTGCGCTGGAGGGTCTCTGG 848
Qy      256  PheGlnAlaGlyIleThrSerPheGlyPheGlyCysGlyArgArgAsnArgProGlyVal 275
Db      849  TACCTGACGGGATCTGTAGCTGGGAGATGCTGTGGGGCCCGCAACAGGCTGTGTG 908
Qy      276  PheThrAlaValAlaThrTyrGluAlaThrTyrGluAlaThrTyrGluAlaMetGlySerGluPro 295
Db      909  TACACTCTGGCTGAGCTATGCTGCTGATCCAAAGCAAGTG-----ACAGAACTC 962
Qy      296  GlyProAlaPheProThrGlnProGlnLysThrGlnSerAsp-----CysLeuHis 312
Db      963  CAGCTTCGTGTGGTGGCCCAACCCAGAGTCCAGCCCGCAGCAGCAACCTCTGTGGCAGC 1022
Qy      313  GlnThrAlaPheLeuAspSer---AlaArgIleLeuLeuArgProLeuSerHisIleSer 331
Db      1023  CACCTGGCTTCAGCTCTGCCCGCAGCCAGGGCTTGTGAGGGCCATCCTTTTCCTGCCT 1082
Qy      332  ValGlyValSerThrGlyThrLysSerLeuValLeuProThrLeuSer----- 347
Db      1083  CTGGGCTGGCTGTGGC-----CTCCTCTCCCATGGCTCAGCGAGCACTGAGCT 1133
Qy      348  -----ProHisSerLeuLeuGlyLeuTrpGlyPhe 357
Db      1134  GGCCCTACTTCCAGGTGATGATCATCATCAAGGACAGGAGCTGTGCTCTTC 1187

RESULT 3
US-09-948-094-1
; Sequence 1, Application US/09948094
; Patent No. US2002090625A1
; GENERAL INFORMATION:
; APPLICANT: Mok, Samuel
; APPLICANT: Wong, Kwong-kwok
; TITLE OF INVENTION: Methods of Detecting Cancer Based on Prostatin
; FILE REFERENCE: 81994/282423
; CURRENT APPLICATION NUMBER: US/09/948,094
; CURRENT FILING DATE: 2001-09-07
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 1
; LENGTH: 1834
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (229)..(1260)
US-09-948-094-1

Alignment Scores:
Pred. No.: 9,11e-54 Length: 1834
Score: 697.00 Matches: 160
Percent Similarity: 55.29% Conservative: 49
Best Local Similarity: 42.33% Mismatches: 133
Query Match: 35.63% Indels: 36
DB: 9 Gaps: 11

US-10-037-417-46 (1-357) x US-09-948-094-1 (1-1834)
Qy      1  MetAlaGlnLysGlyValLeuGlyProGlyGlnLeuGlyAlaValAlaAsnSerAspSer 20
Db      229  ATGGCCCAAGAGGGGTCTCTGGGCTTGGGAGCTGGGGGCTGTGGCC----- 276
Qy      21  TyrSerLeuTyr---GlyLeuValProSerGlyPro-----AlaArgGlyPro 35
Db      277  ATTCTGCTCTATCTTGGATTACTCTCCGTCGGGACAGGAGCGGAGGGGCAAGCTCCC 336
Qy      36  ProTyrCysGlyArgProGluProSerAlaArgIleValGlySerAsnAlaGlnPro 55
Db      337  -----TGCGGT---GTGGCCCCCAAGCAGCATCATCAGGTGGCAGCAGTGTGCGCC 387
Qy      56  GlyThrTrpProThrGlnValSerLeuHisHisGlyGlyGlyHisIleCysGlyGlySer 75

```

```

Db      388  GTCAGTGGCCCTGGCAGGTTCAGATCCTATGAAGCGCTCAGTGTGTGTGGTGGCTCT 447
Qy      76  LeuIleAlaProSerTrpValLeuSerAlaAlaHisCysPheMetThrAsnGlyThrLeu 95
Db      448  CTCGTGTCTGAGCAGTGGTGTCTGTCTGCTGCTCTCTCTCTCTCTCTCTCTCTCTCT 507
Qy      96  GluProAlaAlaGluTrpSerValLeuLeuGlyValHisSerGlnAspGlyProLeuAsp 115
Db      508  GAA-----GCCTATCAGGTCAAGCTGGGGGCCACACAGCTAGCTCTCTCTCTCTCT 558
Qy      116  GlyAlaHisThrArgAlaValAlaAlaIleValProAlaAsnTyrSerGlnValGlu 135
Db      559  GAGCCCAAGGTTCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 618
Qy      136  LeuGlyAlaAspLeuAlaLeuLeuArgLeuAlaSerProAlaSerLeuGlyProAlaVal 155
Db      619  TCCAGGGCGAGTATGACATCTCTCCAACTCAGCAGACCCATCAGCTCTCTCTCTCTCT 678
Qy      156  TrpProValCysLeuProArgAlaSerHisArgPheValHisGlyThrAlaCysTrpAla 175
Db      679  CGGCCCATCTGCTCTCTCTCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 738
Qy      176  ThrGlyTrpGlyAspValGlnGluAlaAspProLeuProLeuProTrpValLeuGlnGlu 195
Db      739  ACTGGCTGGGGTATGATGGCCCTCTCAGTGAGCTCTCTCAGCAGCAGCAGCAGCAGCAG 798
Qy      196  ValGluLeuArgLeuLeuGlyGluAlaThrCysGlnCysLeuTyrSerGlnProGlyPro 215
Db      799  CTCGAGGTGCTCTGATGAGTGTGAGAGCTGTAACTGCTGTACACATCGCAGCGCAAG 858
Qy      216  PheAsnLeuThrLeuGlnIleLeuProGlyMetLeuCysAlaGlyTyrProGluGlyArg 235
Db      859  CCTGAGGAGCGCAGCTTGTCTCAAGAGGACATGCTGTGTCTGTCTGTGTGTGTGTGT 918
Qy      236  ArgAspThrCysGlnGlyAspSerGlyGlyProLeuValCysGluGluGlyGlyArgTrp 255
Db      919  AAGGACGCGCTGCCAGGGTGACTCTGGGGGCCCACTCTCTCTCTCTCTCTCTCTCTCTGG 978
Qy      256  PheGlnAlaGlyIleThrSerPheGlyPheGlyCysGlyArgArgAsnArgProGlyVal 275
Db      979  TACCTGACGGGCATTTGAGCTGGGAGATGCTGTGGGGCCCGCAACAGGCTGTGTGTG 1038
Qy      276  PheThrAlaValAlaThrTyrGluAlaThrTyrGluAlaArgGluGlnValMetGlySerGluPro 295
Db      1039  TACACTGTGGCTCCAGCTATGCTCTCTGATCCAAAGCAAGGAG-----ACAGAACTC 1092
Qy      296  GlyProAlaPheProThrGlnProGlnLysThrGlnSerAsp-----CysLeuHis 312
Db      1093  CAGCTCTGTGTGGTGGCCCAACCCAGGAGTCCAGCCCGCAGCAGCAACCTCTGTGGCAGC 1152
Qy      313  GlnThrAlaPheLeuAspSer---AlaArgIleLeuLeuArgProLeuSerHisIleSer 331
Db      1153  CACCTGGCTCTCAGCTCTGCCCGCAGCCAGGGCTTGTGTGAGGCCCATCTCTTTCTGTGCT 1212
Qy      332  ValGlyValSerThrGlyThrLysSerLeuValLeuProThrLeuSer----- 347
Db      1213  CTGGGCTGTGGCTGTGGC-----CTCCTCTCCCATGGCTCAGCGAGCACTGAGCT 1263
Qy      348  -----ProHisSerLeuLeuGlyLeuTrpGlyPhe 357
Db      1264  GGCCCTACTTCCAGGTGATGATCATCATCAAGGACAGGAGCTGTGCTCTTC 1317

RESULT 4
US-09-880-107-2214
; Sequence 2214, Application US/09880107
; Patent No. US20020142981A1
; GENERAL INFORMATION:
; APPLICANT: Horne, Darci T.
; APPLICANT: Vockley, Joseph G.
; APPLICANT: Scherf, Uwe
; APPLICANT: Gene Logic, Inc.
; TITLE OF INVENTION: Gene Expression Profiles in Liver Cancer

```



```

? FILE REFERENCE: 44921-5028-WO
? CURRENT APPLICATION NUMBER: US/09/880,107
? CURRENT FILING DATE: 2001-06-14
? PRIOR APPLICATION NUMBER: US 60/211,379
? PRIOR FILING DATE: 2000-06-14
? PRIOR APPLICATION NUMBER: US 60/237,054
? PRIOR FILING DATE: 2000-10-02
? NUMBER OF SEQ ID NOS: 3950
? SOFTWARE: PatentIn Ver. 2.1
? SEQ ID NO 2214
? LENGTH: 1834
? TYPE: DNA
? ORGANISM: Homo sapiens
? FEATURE:
? OTHER INFORMATION: Genbank Accession No.
? US-09-880-107-2214

```

Alignment Scores:	
Pred. NO.:	9, 116-54
Score:	697.00
Percent Similarity:	55.23%
Best Local Similarity:	42.33%
Query Match:	35.69%
DP:	9
Length:	1834
Matches:	160
Conservative:	49
Mismatches:	133
Indels:	36
Gaps:	11

1-1834

US-10-037-41748 (1-3577) R 05

QY	1	MetAlaGlnIysGlyValLeuGlyProGlyGlnLeuGlyAlaValAlaAsnSerAspSer	20
Db	229	ATGGCCAGAGGGGGCTCTGGGGCCCTGGGCACCTGGGGCTGGGCC-----	276
QY	21	TyrSerLeuTyr---GlyLeuValProSerGlyPro-----AlaArgGlyPro	35
Db	277	ATTCTGCTCTATCTTGGATTAATCTCCGCTCGGGACAGAGCGAAGGGGCAGAAGCTCCC	336
QY	36	ProTyrCysGlyArgProGluProSerAlaArgIleValGlyGlySerAsnLaGlnPro	55
Db	337	-----TSCGGT-----GTGGCCCCCAAGACGACATCACAGTGCAGCATGCTGCATCGCC	387
QY	56	GlyThrTrpProTrpGlnValSerLeuHisIsglyGlyGlyHisIleCysGlyGlySer	75
Db	388	GGTCAGTGGCCCTGCAGCTAGCATCACTATCAAGGGCTCCCATGTGTGTGGTGGCTCT	447
QY	76	LeuIleAlaProSerTrpValLeuSerAlaAlaHisCysPheMetThrAsnGlyThrLeu	95
Db	448	CTCGTCTCTGACGAGTGGGTGCTGCTCAGTGTCTACTGCTTCCCGACGAGCACCAACAG	507
QY	96	GluProAlaAlaGluTrpSerValLeuLeuGlyValHisSerGlnAspGlyProLeuAsp	115
Db	508	GAA-----GCCATAGAGTCAAGCTGGGGGCCACCACTAGACTCTCTACTCCGAG	558
QY	116	GlyAlaHisThrArgAlaValAlaIleValProAlaAsnTyrSerGlnValGlu	135
Db	559	GAGCCCAAGGTGACGCCCTTGAGGACATCATCCCCACGCCAGCTACTCTCCAGAGGGC	618
QY	136	LeuGlyAlaAspLeuAlaLeuLeuArgLeuAlaSerProAlaSerLeuGlyProAlaVal	155
Db	619	TCCCAGGGCGACATTGCATCTCTCCAATCAGCAGACCCATCAGCTTCTCCCGCTACATC	678
QY	156	TrpProValCysLeuProArgAlaSerHisArgPheValHisGlyThrAlaCysTrpAla	175
Db	679	CGGCCCATCTCCCTCCCTGCAGCCACAGCCCTCTTCCCNACGGCTCCCATGCACTGTC	738
QY	176	ThrGlyTrpGlyAspValGlnGluAlaAspProLeuProLeuProTrpValLeuGlnGlu	195
Db	739	ACTGCTGGGGTCATGTGGCCCCCTCACTGAGCTCTCTGACGCCCAAGCACTGCAGCAA	798
QY	196	ValGluLeuArgLeuLeuGlyGluAlaThrCysGlnCysLeuTyrSerGlnProGlyPro	215
Db	799	CTCAGGTGCTCTCATCAGTCGTGAGACGTGTAACTGCCTGTACACATCGAGCGCAAG	858
QY	216	PheAsnLeuThrLeuGlnIleLeuProGlyMetLeuCysAlaGlyTyrProGluGlyArg	235

859	CCTGAGGAGCGGCATT	TGTC	CAAGAGGACATG	TGTGTG	TGCTGGCTATGTGAGGGGGG	313
236	ArgSerThrCysGlnGlyAspSerGlyGlyProLeuValCysGluGluGlyGlyArgTrp	255				
919	AAGGACGCCTCCCAAGGCTGACTCTG	GGGGGCCAC	TCTCTCTGCCCTGGAGGGTCTCTGG	978		
256	PheGlnAlaGlyIleThrSerPheGlyPheGlyCysGlyArgArgAsnArgProGlyVal	275				
979	TACCTGACGGGCATTGCTAGCTGGGAGATGCCTGTGGGGCCCGAACAGGCTGGTGTG	1038				
276	PheThrAlaValAlaThrTyrGluAlaTrpIleArgGluGlnValMetGlySerGluPro	295				
1039	TACACTCTGGCCTCCAGCTATGCCTCTGGATCCAAAGCAAGGTG	1092				
296	GlyProAlaPheProThrGlnProGlnLysThrGlnSerAsp	312				
1093	CAGCTCGTGTGGTGCCCAAAACCCAGGAGTCCAGCCCGACAGCAACCTCTGTGGCAGC	1152				
313	GlnThrAlaPheLeuAspSer	331				
1153	CACCTGGCCCTTACGCTCTGCCCGCCAGCCGGGTGCTGAGGGCCATCTTTCTCGCT	1212				
332	ValGlyValSerThrGlyThrLysSerLeuValLeuProThrLeuSer	347				
1213	CTGGGCTTGCTCTGGC					
348						
1264	GGCCCTACTTCCAGGATGGATGCATCACTCAAGGACAGGCGCTGGTCTTTC	1317				

RESULT 5

```

US-09-967-768A-141
; Sequence 141, Application US/09967768A
; Patent No. US20020150877A1
; GENERAL INFORMATION:
; APPLICANT: Augustus, Meena
; TITLE OF INVENTION: Cancer
; TITLE OF INVENTION: Sets
; FILE REFERENCE: 689290-72
; CURRENT APPLICATION NUMBER: US/09/967,768A
; CURRENT FILING DATE: 2001-09-28
; PRIOR APPLICATION NUMBER: US/60/236,109
; PRIOR FILING DATE: 2000-09-28
; PRIOR APPLICATION NUMBER: US/60/236,034
; PRIOR FILING DATE: 2000-09-28
; PRIOR APPLICATION NUMBER: US/60/236,111
; PRIOR FILING DATE: 2000-09-28
; NUMBER OF SEQ ID NOS: 325
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 141
; LENGTH: 1834
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-967-768A-141

Alignment Scores:
Pred. No.:          9.11e-54          Length:          1834
Score:              697.00           Matches:         160
Percent Similarity: 55.29%           Conservative:    49
Best Local Similarity: 42.33%        Mismatches:     133
Query Match:        35.69%           Indels:         36
DB:                  9                Gaps:           11

```

1-1834)

US-10-03/-41/-46 (I-557) x 08 02

	Qy	1	MetalGlnInlySglyValLeuGlyProGlyClnLeuGlyAlaValAlaAsnSeraspSer	20
	Dd	229	ATGCCCCAGAAAGGGGTCTCGGCCTGGGCAGCTGGGGGTGTGGCC-----	276
	Qy	21	TyrSerLeuTyr---GlyLeuValProSerGlyPro-----AlaArgGlyPro	35
	Dd	277	ATTCTGTCCTACTTTGGATTACITCCGTGGGGACAGGCGGAAGGGGCAGAAGCTCC	336


```

QY 96 GluProAlaAlaGluTrpSerValLeuLeuGlyValHisSerGlnAspGlyProLeuAsp 115
Db 508 GAA-----GCCATGAGGTCAAGCTGGGGGCCCCACGAGCTAGACTCTCTCTCGAG 558
QY 116 GlyAlaHisThrArgAlaValAlaAlaValValProAlaHisThrSerGlnValGlu 135
Db 559 GAGCCCAAGGTCAACCCCTGAGGAGCATATCCCCACCCAGCTACTCTCCAGGAGGC 618
QY 136 LeuGlyAlaAspLeuAlaLeuLeuAlaSerProAlaSerLeuGlyProAlaVal 155
Db 619 TCCAGGGCGCATGTGACTCTCCAACTCAGCAGACCCATCCTCTCTCCGCTATCATC 678
QY 156 TrpProValCysLeuProArgAlaSerHisArgPheValHisGlyThrAlaCysTrpAla 175
Db 679 CGGCCCATCTGCTCTCTGAGCAGCAACCCCTCTCCCAACAGCGCTCTCCATGCTACTGTC 738
QY 176 ThrGlyTrpGlyAspValGlnGluAlaAspProLeuProLeuProTrpValLeuGlnGlu 195
Db 739 ACTGCTGGGGTCATGTGGCCCTCTCAGTGGAGCTCTCTGAGCCCAAGCCACTGCAGCAA 798
QY 196 ValGlnLeuArgLeuLeuGlyGluAlaThrCysGlnCysLeuTyrSerGlnProGlyPro 215
Db 799 CTCGAGGTGCTCTGATCAGTGTGAGAGCTGTAACTGCTGTACACATCGACGCCCAAG 858
QY 216 PheAsnLeuThrLeuGlnLeuLeuProGlyMetLeuCysAlaGlyTyrProGluGlyArg 235
Db 859 CCTGAGGAGCGCACTTTGTCTCAAGAGGACATGGTGTGTGCTGTGCTGTGAGGGGGGC 918
QY 236 ArgAspThrCysGlnGlyAspSerGlyGlyProLeuValCysGluGluGlyGlyArgTrp 255
Db 919 AAGGAGCGCTGCGAGGTGATCTGCGGGCCCACTCTCTGCTGCTGCTGAGGGGTCTCTGG 978
QY 256 PheGlnAlaGlyLeuThrSerPheGlyPheGlyCysGlyArgArgAsnArgProGlyVal 275
Db 979 TACTGAGGGGCACTGTGAGCTGGGAGATGCTGTGGGGCCCGCAACAGCGCTGTGTG 1038
QY 276 PheThrAlaValAlaThrGlyAlaTrpIleArgGluGlnValMetGlySerGluPro 295
Db 1039 TACACTTGGCTCCAGCTATGCTCTCTGATCCAAAGAGGTG-----ACAGAATC 1092
QY 296 GlyProAlaPheProThrGlnProGlnLysThrGlnSerAsp-----CysLeuHis 312
Db 1093 CAGCTCTGTGTGGTGGCCCAACCCAGGAGTCCCGAGCCGACCAACCTCTGTGGCAGC 1152
QY 313 GlnThrAlaPheLeuAspSer---AlaArgIleLeuLeuArgProLeuSerHisIleSer 331
Db 1153 CACCTGGCCCTTCAGCTGCGCCCAAGCCAGCGGCTGTGAGGCGCATCTTTCTGCT 1212
QY 332 ValGlyValSerThrGlyThrLysSerLeuValLeuProTrpLeuSer----- 347
Db 1213 CTGGGCTGTGCTGGGC-----CTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1263
QY 348 -----ProHisSerLeuLeuGlyLeuTrpGlyPhe 357
Db 1264 GGCCTTACTTCCAGATGGATGCATCAGCTCAAGCAGAGGAGCGCTGTCTCTCTCT 1317

RESULT 7
US-10-101-510-447
; Sequence 447, Application US/10101510
; Publication No. US20030148295A1
; GENERAL INFORMATION:
; APPLICANT: WANG, YIXIN
; TITLE OF INVENTION: EXPRESSION PROFILES AND METHODS OF USE
; FILE REFERENCE: 15117.0012
; CURRENT APPLICATION NUMBER: US/10/101,510
; PRIOR FILING DATE: 2002-03-20
; PRIOR FILING DATE: 2001-03-20
; NUMBER OF SEQ ID NOS: 805
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 447
; LENGTH: 3382

```

```

; TYPE: DNA
; ORGANISM: Homo sapiens
; US-10-101-510-447

Alignment Scores: 1,76e-53 Length: 3382
Pred. No.: 697.00 Matches: 160
Score: 55.29% Conservative: 49
Best Similarity: 42.33% Mismatches: 133
Best Local Similarity: 35.69% Indels: 36
Query Match: 14 Gaps: 11
DB:

US-10-037-417-46 (1-357) x US-10-101-510-447 (1-3382)

QY 1 MetAlaGlnLysGlyValLeuGlyProGlyGlnLeuGlyAlaValAlaAsnSerAspSer 20
Db 230 ATGGCCCAAGAGGGTCTCTGGGGCTGGGAGCTGGGGGCTGTGGCC----- 277
QY 21 TyrSerLeuTyr---GlyLeuValProSerGlyPro-----AlaArgGlyPro 35
Db 278 ATTCTGCTCTATCTTGGATTACTCCGGTCCGGGACAGGAGCGGAAGGGCAGAAAGCTCCC 337
QY 36 ProTyrCysGlyArgProGluProSerAlaArgIleValGlyGlySerAsnAlaGlnPro 55
Db 338 -----TGCAGT---GTGGCCCCCAAGCAGCGCATCAGAGTGGCAGCAGTCCGCC 388
QY 56 GlyThrTrpProTrpGlnValSerLeuHisIleGlyGlyHisIleCysGlyGlySer 75
Db 389 GGTGAGTGGCCCTGGCAGTCCAGCATCACCTATGAGGGCTCCATGTGTGTGTGGTCTCT 448
QY 76 LeuIleAlaProSerTrpValLeuSerAlaAlaHisCysPheMetThrAsnGlyThrLeu 95
Db 449 CTCGTGTCTGACGACGTGGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 508
QY 96 GluProAlaAlaGluTrpSerValLeuLeuGlyValHisSerGlnAspGlyProLeuAsp 115
Db 509 GAA-----GCCTATGAGGTCAAGCTGGGGGCCCAACAGCTAGCTAGCTTCTTCTCC 559
QY 116 GlyAlaHisThrArgAlaValAlaAlaIleValValProAlaAsnTyrSerGlnValGlu 135
Db 560 GAGCCCAAGGTCAAGCTGAGGAGCATATCCCCACCCAGCTACTCTCCAGGAGGGC 619
QY 136 LeuGlyAlaAspLeuAlaLeuLeuArgLeuAlaSerProAlaSerLeuGlyProAlaVal 155
Db 620 TCCAGGGCGCATGTGACTCTCTCCAACTCAGCAGACCCATCATCTCTCTCCGCTATCAT 679
QY 156 TrpProValCysLeuProArgAlaSerHisArgPheValHisGlyThrAlaCysTrpAla 175
Db 680 CGGCCCATCTGCTCTCTCTGAGCAGCAAGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 739
QY 176 ThrGlyTrpGlyAspValGlnGluAlaAspProLeuProLeuProTrpValLeuGlnGlu 195
Db 740 ACTGGCTGGGGTCTATGTGGGGCCCTCAGTGGAGCTCTCTGAGCCCAAGCCACTGCAGCA 799
QY 196 ValGlnLeuArgLeuLeuGlyGluAlaThrCysGlnCysLeuTyrSerGlnProGlyPro 215
Db 800 CTCGAGGTGCTCTGATCAGTGTGAGCAGTGTAACTGCCTGTACAACTCAGCAGCCCAAG 859
QY 216 PheAsnLeuThrLeuGlnIleLeuProGlyMetLeuCysAlaGlyTyrProGluGlyArg 235
Db 860 CCGAGGAGCGGCATTTGTCTCAAGAGGACATGTGTGTGTGTGTGTGTGTGTGTGTGTGT 919
QY 236 ArgAspThrCysGlnGlyAspSerGlyGlyProLeuValCysGluGluGlyGlyArgTrp 255
Db 920 AAGGAGCGCTGCCAGGTGATCTCTGGGGCCCACTCTCTCTCTCTCTCTCTCTCTCTCTCT 979
QY 256 PheGlnAlaGlyIleThrSerPheGlyPheGlyCysGlyArgArgAsnArgProGlyVal 275
Db 980 TACCTGAGGGCATTTGTGAGCTGGGAGATGCTCTGTGGGGCCCGCAACAGGCTGTGTG 1039
QY 276 PheThrAlaValAlaThrGlyAlaTrpIleArgGluGlnValMetGlySerGluPro 295
Db 1040 TACACTTGGCCCTCCAGCTATGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1093
;

```


388 GGTCACTGGCCCTGGCAGGTCACATCACCTATGAAGCGGTCCATGTGTGTGGTGGCTCT 447
DB

76 Leu1AalProSerTrpValLeuSerAlaAlahisCysPheMetThrAsnGlyThrLeu 95
QY |||:::|||||
DB CTCGTGTCTGCAGTGCGGTGCTGTACGTCTCACTGCTTC-----489

96 GluProAlaAlaGluTrpSerValLeuLeuGlyValHisSerGlnAspGlyProLeuAsp 115
QY |||:::|||||
DB -----489

116 GlyAlaHisThrArgAlaValAlaAlaValValProAlaAsnTyrsSerGlnValGlu 135
QY |||:::|||||
DB -----CCCAGGCGACACCACAGGCGCTCC 513

136 LeuGlyAlaaspLeuAlaLeuArgLeuAlaSerProAlaSerLeuGlyProAlaVal 155
QY |||:::|||||
DB CAGGCG---GACATTGCCACTCTCTCACTCAGCAGACCCACCAGCTACTCCCGCTACATC 570

156 TrpProValCysLeuProArgAlaSerHisArgPheValHisGlyThrAlaCysTrpAla 175
QY |||:::|||||
DB CGGCCCATCTGCTCCCTGCAGCAACGCTCTCTCCCAACAGGCGCTCCCACTGCACCTGTC 630

176 ThrGlyTrpGlyAspValGlnGluAlaaspProLeuProLeuProTrpValLeuGlnGlu 195
QY |||:::|||||
DB ACTGGCTGGGTCATGTGGCCCCCTCAGTAGGCTCTCAGCGCCCAAGCCACTGCAGCAA 690

196 ValGluLeuArgLeuLeuGlyGluAlaThrCysGlnCysLeuTyrsSerGlnProGlyPro 215
QY |||:::|||||
DB CTGAGGTCGCTCTGTATCAGTCTGTGAGACGTGTAACTGCTGTACAACTCAGCGCCAAG 750

216 PheAsnLeuThrLeuGlnLeuLeuProGlyMetLeuCysAlaGlyTyrsProGluGlyArg 235
QY |||:::|||||
DB CCTGAGGCGCGCACTTTGTCCAAAGGACATGCTGTCTGTCTGTATGTGAGGGGGGC 810

236 ArgAspThrCysGlnGlyAspSerGlyGlyProLeuValCysGluGluGlyGlyArgTrp 255
QY |||:::|||||
DB AAGCAGGCTCGCCAGGCTGACTCTGGGGAACCACTCTCTCTGCTGTGAGGGTCTCTGG 870

256 PheGlnAlaGlyIleThrSerPheGlyPheGlyCysGlyArgArgAsnArgProGlyVal 275
QY |||:::|||||
DB TACCTGACGGGCATTTGAGCTGGGAGATGCTCTGGGGCCGCAACAGGCGCTGGTGG 930

276 PheThrAlaValAlaThrTyrgluAlaTrpIleArgGluGlnValMetGlySerGluPro 295
QY |||:::|||||
DB TACACTCTGGCTCCAGCTATGCTCTCTGTGATCCAAAGCAAGGTG-----ACAGNACTC 984

296 GlyProAlaPheProThrGlnProGlnTyrsThrGlnSerAsp-----CysLeuHis 312
QY |||:::|||||
DB CAGCCTCTGTGTGGTGGCCCAACCCAGGAGTCCCGCCGACAGCAACCTCTGTGGCAGC 1044

313 GlnThrAlaPheLeuAspSer---AlaArgIleLeuLeuArgProLeuSerHisIleSer 331
QY |||:::|||||
DB CACCTGGGCTTCAGCTCTGCCCGCCAGGCGTTGCTCAGGCGCCATCTCTTCTCGCT 1104

332 ValGlyValSerThrGlyThrLysSerLeuValLeuProTrpLeuSer-----347
QY |||:::|||||
DB CTCTCTCTCCCATGGTTCAGCGAGCTAGAGCTAGCTAGCTAGCTAGCTAGCTAGCT 1155

348 -----ProHisSerLeuLeuGlyLeuTrpGlyPhe 357
QY |||:::|||||
DB GGCCCTACTCCAGGTGATGATGATCACACTCAAGGACAGGCGCTGGTCTCTC 1209

RESULT 10
US-10-109-616-1
; Sequence 1, Application US/10109616
; Publication No. US20030167484A1
; GENERAL INFORMATION:
; APPLICANT: Allen, Keith D.
; TITLE OF INVENTION: TRANSGENIC MICE CONTAINING CHANNEL
; TITLE OF INVENTION: ACTIVATING PROTEASE 1 (CAP1) GENE DISRUPTIONS
; FILE REFERENCE: R-490
; CURRENT APPLICATION NUMBER: US/10/109,616
; CURRENT FILING DATE: 2002-03-28

; PRIOR APPLICATION NUMBER: US 60/280,509
 ; PRIOR FILING DATE: 2001-03-29
 ; PRIOR APPLICATION NUMBER: US 60/311,055
 ; PRIOR FILING DATE: 2001-08-08
 ; NUMBER OF SEQ ID NOS: 4
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 1
 ; LENGTH: 1797
 ; TYPE: DNA
 ; ORGANISM: Mus musculus
 US-10-109-616-1

Alignment Scores:

Pred. No.: 1,72e-47 Length: 1797
 Score: 627.50 Matches: 144
 Percent Similarity: 55.17% Conservatives: 48
 Best Local Similarity: 41.38% Mismatches: 139
 Query Match: 32.13% Indels: 17
 DB: Gaps: 9

US-10-037-417-46 (1-357) x US-10-109-616-1 (1-1797)

QY 1 MetAlaGlnysGlyValLeuGlyProGlyGlnLeuGlyAlaValAlaAenSerAspSer 20
 DB 146 ATGCCCCAAGGTGGGCTGGGACTTTGGGAGCTGGAAAGCTGTGACC-----ATT 196
 QY 21 TyrSerLeuTyrGlyLeuValProSerGly---ProAlaargGly---ProProTyrCys 38
 DB 197 CTGCTCTCTCGGATTCCTCCAGTCCGGAATCCGAGTACGAGCTGACGGACTCGTCTCTGT 256
 QY 39 GlyArgPro---GluProSerAlaArgIleValGlyGlySerAsnAlaGlnProGlyThr 57
 DB 257 GGTCCCTTCATCCAGCCA-----CGATCACCGTGTGTGCGAGTGCAGGCCCGGTGAG 310
 QY 58 TrpProTyrGlnValSerLeuHisGlyGlyGlyHisGlyCysGlyGlySerLeuIle 77
 DB 311 TGGCCCTCGGAGTGCAGCATCACCTAGTATGCGCAACCATGTTTGTGGCGGTCTCTGTG 370
 QY 78 AlaProSerTrpValLeuSerAlaAlaHisCysPheMetThrAsnGlyThrLeuGluPro 97
 DB 371 TCAATAAATGGGTGGTGTCTGCTGCTACACGCTTCCCGAGAGAACACAGCGAGAA--- 427
 QY 98 AlaAlaGluTrpSerValLeuGlyValHisSerGlnAspGlyProLeuAspGlyAla 117
 DB 428 -----GGTATGAGTGAAGCTGGGGCCACCAGTACTAGTCTCTACAGCAATGACACT 481
 QY 118 HisThrArgAlaValAlaAlaIleValProAlaAenTyrSerGlnValGluLeuGly 137
 DB 482 GTGGTCCACAGTGGCTCAGATCATCACCCACTCAAGCTACCGAGAGAGGGTCCCGAG 541
 QY 138 AlaAspLeuAlaLeuLeuArgLeuSerProAlaSerLeuGlyProAlaValTrpPro 157
 DB 542 GGGGACATCGGTTCTCCGCTCAGCAGTCTGTCTCTCTCCGAGTACATCAGACCC 601
 QY 158 ValCysLeuProArgAlaSerHisArgPheValHisGlyThrAlaCysTrpAlaThrGly 177
 DB 602 ATTGCTCTCCGAGCAATAGCTCTCTCCAGCGGCTTCACTGACTGTACAGGGA 661
 QY 178 TrpGlyAspValGlnGluAlaAspProLeuProLeuProTrpValLeuGlnGluValGlu 197
 DB 662 TGGGGTCTATGGCTCTCTCAGTACGAGCTCCAGACCCCTAGCGCTCTCGAGCAGTCCAG 721
 QY 198 LeuArgLeuLeuGlyGluAlaThrCysGlnCysLeuTyrSerGlnProGlyProPheAsn 217
 DB 722 GTACCATCATCGCGGGAACCTGTAGTCTGCTGCTACCAATTAATGCGGTGCTGAA 781
 QY 218 LeuThrLeuGlnIleLeuProGlyMetLeuCysAlaGlyTyrProGluGlyArgAsp 237
 DB 782 GAACCGCACACTATCCAGCAGACATGCTGTGTGTGCTATGTGAGGAGGAGGAGGAT 841
 QY 238 ThrCysGlnGlyAspSerGlyGlyProLeuValCysGluGluGlyGlyArgTrpPheGln 257
 DB 842 GCCTGCGAGGTGACTCTGGGGGCCCACTCTCTTGTCCAGGAGGAGGAGGAGTCTGTTG 901

RESULT 11

US-10-042-865-31
 ; Sequence 31, Application US/10042865
 ; Publication NO. US20040029216A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Padigar, Muralidhara
 ; APPLICANT: Li, Li
 ; APPLICANT: Zerhusen, Bryan D
 ; APPLICANT: Casman, Stacie J
 ; APPLICANT: Shenoy, Suresh G
 ; APPLICANT: Spytek, Kimberly
 ; APPLICANT: Zhong, Mei
 ; APPLICANT: Gangolli, Esha A
 ; APPLICANT: Burgess, Catherine E
 ; APPLICANT: Patturajan, Meera
 ; APPLICANT: Vernet, Corine A.M
 ; APPLICANT: Taylor, Sarah
 ; APPLICANT: Tchernev, Velizar T
 ; APPLICANT: Miller, Charles E
 ; APPLICANT: Guo, Xiaojia
 ; APPLICANT: Boldog, Ference L
 ; APPLICANT: Grosse, William M
 ; APPLICANT: Alsbrook II, John P
 ; APPLICANT: Gerlach, Valerie L
 ; APPLICANT: Edinger, Shlomit R
 ; APPLICANT: Rothenberg, Mark E
 ; APPLICANT: Ellerman, Karen
 ; APPLICANT: Macdougall, John
 ; APPLICANT: Malyankar, Uriel M
 ; APPLICANT: Millet, Isabelle
 ; APPLICANT: Peyman, John
 ; APPLICANT: Smithson, Glennda
 ; APPLICANT: Gunther, Erik
 ; APPLICANT: Stone, David
 ; TITLE OF INVENTION: Proteins, Polynucleotides Encoding Them and Methods of
 ; FILE OF INVENTION: Using the Same
 ; FILE REFERENCE: 21402-537
 ; CURRENT APPLICATION NUMBER: US/10/042,865
 ; CURRENT FILING DATE: 2002-05-17
 ; PRIOR APPLICATION NUMBER: 60/260,417
 ; PRIOR FILING DATE: 2001-01-09
 ; PRIOR APPLICATION NUMBER: 60/260,831
 ; PRIOR FILING DATE: 2001-01-10
 ; PRIOR APPLICATION NUMBER: 60/272,338
 ; PRIOR FILING DATE: 2001-02-28
 ; PRIOR APPLICATION NUMBER: 60/274,876
 ; PRIOR FILING DATE: 2001-03-09
 ; PRIOR APPLICATION NUMBER: 60/284,704
 ; PRIOR FILING DATE: 2001-04-18
 ; NUMBER OF SEQ ID NOS: 264
 ; SOFTWARE: PatentIn Ver. 2.1
 ; SEQ ID NO 31

QY 258 AlaGlyIleThrSerPheGlyPheGlyCysGlyArgArgAsnArgProGlyValPheThr 277
 DB 902 GCAGGCATTTGAGTTGGGTGATGCCCTGTGTGCCCCAACAGGCTGGAGTATACACT 961
 QY 278 AlaValAlaThrTyrGluAlaTrpIleArgGluGlnValMetGlySerGluProGlyPro 297
 DB 962 CTGACTTCTACCTATGCTTCTCTGGATCCACCATGTG-----GCAGAGCTCCAGCCT 1015
 QY 298 AlaPheProThrGlnProGlnLysThrGlnSerAsp-----CysLeuHisGlnThr 314
 DB 1016 CGAGTTGTCCCCGAGACTCAGGAGTCCCGAGCGGATGGCCACTCTGACCATCATCT 1075
 QY 315 AlaPheLeuAspSerAlaArg---IleLeuLeuArgProLeuSerHisIleSerValGly 333
 DB 1076 GTCTTCAGCTCAGCGGCGAGCCCGAAATTTGTAAGCCCGTACTTTTCTGCCACTTGGT 1135
 QY 334 ValSerThrGlyThrLysSerLeu 341
 DB 1136 CTGACCTGGGCTCTTGTCTTTG 1159

```
; LENGTH: 1161
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-042-865-31

Alignment Scores:
Pred. No.:      2,7e-46      Length:      1161
Score:          612.00      Matches:      134
Percent Similarity: 48.38%      Conservative: 45
Best Local Similarity: 36.22%      Mismatches: 121
Query Match:      31.34%      Indels:      70
DB:              12          Gaps:      5

US-10-037-417-46 (1-357) x US-10-042-865-31 (1-1161)
QY 1 MetAlaGlnLysGlyValLeuGlyProGlyGlnLeuGlyAlaValAla-----16
Db 1 ATGGCCAGAGAGGGGCTCTGGGGCTGGGAGTGGGGCTGGGCGCATCTGCTCTAT 60
QY 17 -----AsnSerAspSerTyrSerLeuTyrGly 25
Db 61 CTGTGATTACTCCGGTCGGGACAGGAGCGGAGGAGGAGGAATGCCAGATTGGT 120
QY 26 LeuVal-----27
Db 121 GTGGTGGAAACAGACACAGATTGTGACTCCGGAACAATTGTGAGACTGGGCGCATCGGG 180
QY 28 -----ProSerGlyProAlaArg 33
Db 181 GAGCATGTTGGAGGTTGGGGGGTGTGGGCTCGGAGCAATGCCCACTCAGCCGACAG 240
QY 34 GlyProProTyrCysGly-----ArgProGlu-----42
Db 241 AGCCCGCCTGTCTACGAGGGGCTGTGTTCATCAAGCATCCACAGCAAGGAAGT 300
QY 43 -----ProSerAlaArgIleValGlyGlySerAsnAla 53
Db 301 TTGCCAGACACTCCCTGCGGTGTGGCCCGCCCAAGCAGCATCACAGGTGGCAGTGCA 360
QY 54 GlnProGlyThrTrpProTyrGlnValSerLeuHisGlyGlyGlyHisIleCysGly 73
Db 361 GTGCGCGGTTCAGTGGCCCTGGGCGGTCAGTGCATCCCTATGAAGGCGCTCATGTGTGT 420
QY 74 GlySerLeuIleAlaProSerTrpValLeuSerAlaAlaHisCysPheMetThrAsnGly 93
Db 421 GGTCTCTCTGTGCTGAGCAGTGGTGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 480
QY 94 ThrLeuGluProAlaAlaGluTrpSerValLeuLeuGlyValHisSerGlnAspGlyPro 113
Db 481 CACAAGGAA-----GCTATGAGGTCAAGCTGGGGGCCCCACCACTAGACTCCTTAC 531
QY 114 LeuAspGlyAlaHisThrArgAlaValAlaAlaIleValValProAlaAsnTyrSerGln 133
Db 532 TCCGAGGAGCCCAAGGTACAGACCCCTGAAGGAGCATATCCCCACCCAGCTACTCCAG 591
QY 134 ValGluLeuGlyAlaAspLeuAlaLeuLeuArgLeuAlaSerProAlaSerLeuGlyPro 153
Db 592 GAGGGTCCAGGGGACATTTGACTCTCTCAACTCAGCAGAGCCCATCACTTCTCCCGC 651
QY 154 AlaValTrpProValCysLeuProArgAlaSerHisArgPheValHisGlyThrAlaCys 173
Db 652 TACATCGGCCCATCTGCTCCCTGAGCCCAAGCGCTCTTCCCAACAGCGCTCCACTGC 711
QY 174 TrpAlaThrGlyTrpGlyAspValGlnGluAlaAspProLeuProLeuTrpValLeu 193
Db 712 ACTGTCACTGGCTGGGTTCATGTGGCCCTCTCATGTAGGCTCTCTGAGCGCCCAAGCCACTG 771
QY 194 GlnGlnValGluLeuArgLeuLeuGlyGluAlaThrCysGlnCysLeuTyrSerGlnPro 213
Db 772 CAGCACTCGAGGTGCCTCTGATCAGTGTGAGAGCGTAACTGCTGTGTACAACTCGAC 831
QY 214 GlyProPheAsnLeuThrLeuGlnIleLeuProGlyMetLeuCysAlaGlyTyrProGlu 233
Db -----
```

```
832 GCCAAGCCTGAGGAGCGGCACTTTGTCCAGAGGACATGGTGTGTGCTGCTATGTGGAG 891
234 GlyArgArgAspThrCysGlnGlyAspSerGlyGlyProLeuValCysGlnGluGlyGly 253
892 GGGGCGAAGGAGCGCTGCCAGGGTGACTCTGGGGGCCCACTCTCTCCCTGTGGAGGGT 951
254 ArgTrpPheGlnAlaGlyIleThrSerPheGlyPheGlyCysGlyArgArgAsnArgPro 273
952 CTCTGGTACCTGACGGGCATTTGAGCTGGGGAGATGCCCTGTGGGGCCCGCAACAGGCCT 1011
274 GlyValPheThrAlaValAlaThrTyrGluAlaTrpIleArgGlnValMetGlySer 293
1012 GGTGTGTACTCTGGGCTCCAGTATGCTCTGGATCCAAAGCAAGGACTCTCTCGGGA 1071
294 GluProGlyProAlaPheProThrGlnPro 303
1072 GGCTGGGGCCCATCTTGATCTTTGAGCCC 1101

RESULT 12
US-10-041-400A-1
; Sequence 1, Application US/10041400A
; Publication No. US2002110895A1
; GENERAL INFORMATION:
; APPLICANT: Darrow, Andrew
; APPLICANT: Andrade-Gordon, Patricia
; APPLICANT: Qi, Jensen
; TITLE OF INVENTION: DNA Encoding the Human Serine
; TITLE OF INVENTION: Protease EOS
; FILE REFERENCE: ORT-1031
; CURRENT APPLICATION NUMBER: US/10/041,400A
; CURRENT FILING DATE: 2002-01-08
; PRIOR APPLICATION NUMBER: US/09/387,375
; PRIOR FILING DATE: 1999-08-31
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 1
; LENGTH: 1613
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-041-400A-1

Alignment Scores:
Pred. No.:      2,26e-45      Length:      1613
Score:          603.50      Matches:      139
Percent Similarity: 50.00%      Conservative: 38
Best Local Similarity: 39.27%      Mismatches: 129
Query Match:      30.90%      Indels:      48
DB:              13          Gaps:      7

US-10-037-417-46 (1-357) x US-10-041-400A-1 (1-1613)
QY 38 CysGlyArgProGluProSerAlaArgIleValGlyGlySerAsnAlaGlnProGlyThr 57
Db 150 TCGGGCGAGCCCGCATGTCTCCAGTCCGATGTTGGGGCGCGGATGGCGGAGCGGAGAG 209
QY 58 TrpProTrpGluValSerLeuHisGlyGlyGlyHisIleCysGlyGlySerLeuIle 77
Db 210 TGGCGGTGGCAGCGGAGCATCTCAGATCTCTGGGGCACAGTGTGGGGGGTGGCTCATC 269
QY 78 AlaProSerTrpValLeuSerAlaAlaHisCysPheMetThrAsnGlyThrLeuGluPro 97
Db 270 GCCCCCCAGTGGGTGTGACAGCGGCGCATGTCTTCCCCAGGAGGCGACTG-----320
QY 98 AlaAlaGluTrpSerValLeuLeuGlyValHisSerGlnAspGlyProLeuAspGlyAla 117
Db 321 CCAGCTGAGTACCGCGCTGGCGCTGGGGCGGCTGCTGCTGCTCCACCTCGCCCGCCACG 380
QY 118 HisThrArgAlaValAlaAlaIleValValProAlaAsnTyrSerGlnValGluLeuGly 137
Db 381 CTCTCGGTGGCGGCGGAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 440
QY 138 AlaAspLeuAlaLeuLeuArgLeuAlaSerProAlaSerLeuGlyProAlaValTrpPro 157
Db -----
```

441 GGCGAGCTGGCACTGCTGCAGCTGGGTGCGCCGGTGCCTCGAGCGCTCGCGTCCACCC 500

158 ValCysLeuProArgAlaSerHisArgPheValHisGlyThrAlaCysTrpAlaThrGly 177

501 GTCGTGCTGCGGTGCGCGCGCGCGCGCGCGCGCGCGCACACCATGCGGGTCAACGGC 560

178 TrpGlyAspValGlnGluAlaAspProLeuProLeuProLeuProTrpValLeuGlnValGlu 197

561 TGGGGGAGCTCCGCGCCAGGAGTGCCTCCAGAGTGGCGACCGCTACAAGAGTAAGG 620

198 LeuArgLeuLeuGlyGluAlaThrCysGlnCysLeuTyr-----SerGlnPro 213

621 GTGCCCTGCTGACCTGCGCACCTCGAGCGGCTCTTACAGTGGCGCGGACGTGCCCC 680

214 GlyProPheAsnLeuThrLeuGlnIleLeuProGlyMetLeuCysAlaGlyTyrProGlu 233

681 CAGGCTGAGGCATT-----GTGCTGCTGGGAGTCTGTGTGCGGCTACCCCGCAG 731

234 GlyArgAspThrCysGlnGlyAspSerGlyGlyProLeuValCysGluGluGlyGly 253

732 GGCCACAAGACGCCCTGCCAGGGTGATTTCTGGGGGACCTCTGACCTGCTGCGAGTCTGGG 791

254 ArgTrpPheGlnAlaGlyIleThrSerPheGlyPheGlyCysGlyArgAsgAsnArgPro 273

792 AGCTGGGTCTGTTGGCGGTGGTGGAGCTGGGGCAAGGGTTGTGCCCTGCCCAACCGTCCA 851

274 GlyValPheThrAlaValAlaThrTyrGluAlaTrpIleArgGluGlnValMetGlySer 293

852 GGGGTCTACACCAAGTGTGGCCACATATAGCCCTGGATTGAGGCTCGCGTCACTTCTAAT 911

294 Glu-----ProGly----- 296

912 GCTAGCCGGTGAAGCTGACCTGGAGCCAGCTGTGGGGTCCCTCAGCTCTCGTTTCATC 971

297 -----ProAlaPhe-ProThrGlnProGlnLysThrGlnSerAspCysLeuHisGlnTh 314

972 CAGGCACCTGCCTATACCCACATCCCTCTTGCTCGAGGCCAAGATGCTATAAAAG-- 1029

314 rAlaPheLeuAspSerAlaArgIleLeuLeuArgProLeuSerHisIleSerValGlyVa 334

1030 -----CTAAAGGCCGCCACCCACCCACCTTCTGGCTCTCT 1073

334 lSerThrGlyThrIysSerLeuValLeuProTrpLeuSer----- 347

1074 CCTCTTTGGGATCACAGCTCTGACTTCACCAACCTTCACAGGAATCTGCCATGAGT 1133

348 -----ProHisSerLeuGlyLeuTrp 355

1134 CCCAGGAGTCAACACTCCCCACTCCCTTCTGGCTTGAT 1173

RESULT 13

US-10-041-264A-1

Sequence 1, Application US/10041264A

Publication No. US2002012446A1

GENERAL INFORMATION:

APPLICANT: Darrow, Andrew

APPLICANT: Andrade-Gordon, Patricia

APPLICANT: Qi, Jensen

TITLE OF INVENTION: DNA Encoding the Human Serine

TITLE OF INVENTION: Protease EOS

FILE REFERENCE: ORT-1031

CURRENT APPLICATION NUMBER: US/10/041,264A

CURRENT FILING DATE: 2002-01-08

PRIOR APPLICATION NUMBER: US/09/387,375

PRIOR FILING DATE: 1999-08-31

NUMBER OF SEQ ID NOS: 9

SOFTWARE: PatentIn Ver. 2.0

SEQ ID NO 1

LENGTH: 1613

TYPE: DNA

ORGANISM: Homo sapiens

US-10-041-264A-1

Alignment Scores:			
Pred. No.:	2,26e-45	Length:	1613
Score:	603.50	Matches:	139
Percent Similarity:	50.00%	Conservative:	38
Best Local Similarity:	39.27%	Mismatches:	129
Query Match:	30.90%	Indels:	48
DB:	13	Gaps:	7
US-10-037-417-46 (1-357) x US-10-041-264A-1 (1-1613)			
QY	38	CysGlyArgProGluProSerAlaArgIleValGlyGlySerAsnAlaGlnProGlyThr	57
DB	150	TGCGGGCAGCCCGCATGCTCCAGTCCGATCGTTGGGGCGCGGATGGCCGCGACGAGAG	209
QY	58	TrpProTndlnValSerLeuHisHISGlyGlyGlyHisIleCysGlyGlySerLeuIle	77
DB	210	TGCGCGTGGCAGCGAGCATCCAGCATCTCTGGGGCACACGTGTGCGGGGGTTCGCTCATC	269
QY	78	AlaProSerTrpValLeuSerAlaAlaHisCysPheMetThrAsnGlyThrLeuGluPro	97
DB	270	GCCTCCAGTGGGTGGTGTGACAGCGGCACATGCTTCTCCCGAGGAGGCACTG-----	320
QY	98	AlaAlaGluTrpSerValLeuLeuGlyValHisSerGlnAspGlyProLeuAspGlyAla	117
DB	321	CCAGCTCAGTACCGCGTGCCTCTGGGGGGGCTGCGTCTGGGCTCCACCTTCGCCCGCACG	380
QY	118	HisThrArgAlaValAlaAlaIleValProAlaAsnTrpSerGlnValGluLeuGly	137
DB	381	CTCTCGGTGCGGTGCGACCGGGTGTGCTGCCCGCGACTACTCCGAGGACGGGGCCCGC	440
QY	138	AlaAspLeuAlaLeuLeuArgLeuAlaSerProAlaSerLeuGlyProAlaValTrpPro	157
DB	441	GGCGACCTGGCAGCTGTCTGAGCTCGCTGCGCGCGTGCCTTGAGCGCTGCGCTCCAAACC	500
QY	158	ValCysLeuProArgAlaSerHisArgPheValHisGlyThrAlaCysTrpAlaThrGly	177
DB	501	GTCTGCTCTGCCGTGCCCGGGCGCCGCCCGCGCGCGGCACACCTTCCGGGTACCGGC	560
QY	178	TrpGlyAspValGlnGluAlaAspProLeuProLeuProTrpValLeuGlnGluValGlu	197
DB	561	TGGGGCAGCTCCGCCCGCAGGAGTGCCCTCCAGAGTGGCGACCGCTACAAAGAGTAAGG	620
QY	198	LeuArgLeuLeuGlyGluAlaThrCysGlnCysLeuTyr-----SerGlnPro	213
DB	621	GTGCGCTGTCTGACTCCGCGACCTCGCAGCGGCTCTTACCACGTGGCGCGGACGTGCC	680
QY	214	GlyProPheAsnLeuThrLeuGlnIleLeuProGlyMetLeuCysAlaGlyTyrProGlu	233
DB	681	CAGGCTGAGCGCAAT-----GTGCTGCTGGGAGTCTGTGTGCGGCTACCCCGCAG	731
QY	234	GlyArgArgAspThrCysGlnGlyAspSerGlyGlyProLeuValCysGluGluGlyGly	253
DB	732	GGCCACAGGACGCGTGCAGGGTGATCTCTGGGGGACCTCTGACCTCCCTGCGACTGTGG	791
QY	254	ArgTrpPheGlnAlaGlyIleThrSerPheGlyPheGlyCysGlyAspArgAsnArgPro	273
DB	792	AGCTGGGTCTGTGGGGGTGGTGGAGCTGGGGCAAGGGTGTGGCTTGCCTGCCCAACCTTCCA	851
QY	274	GlyValPheThrAlaValAlaThrTyrGluAlaTrpIleArgGluGlnValMetGlySer	293
DB	852	GGGGTCTACACCAAGTGTGGCCACATATAGCCCTTGATTCAGGCTCGCGTCACTTCTAAT	911
QY	294	Glu-----ProGly-----	296
DB	912	GCTAGCCCGTGAAGCTGACCTGGAGCCAGCTGTGGGGTCCCTCAGGCTCTCTGGTTCACTC	971
QY	297	-----ProAlaPhe-ProThrGlnProGlnIleThrGlnSerAspCysLeuHisGlnThr	314
DB	972	CAGGCACCTGTCTATACCCACATCCCTTCTGCTCGAGGCCAAGATGCTCTTAAAAAG--	1029
QY	314	rAlaPheLeuAspSerAlaArgIleLeuLeuArgProLeuSerHisIleSerValGlyVal	334
DB	1030	-----CTAAAGGCCACCCACCCCGCCACCCACCTCTTGGCTCTCTCT	1073


```

QY 334 lserThrGlyThrLysSerLeuValLeuProTrpLeuSer----- 347
Db 1074 CCTCTTTGGGGATACACAGCTCTGACTCCACCAACCTCTCAGGAATCTGCCATGAGT 1133
QY 348 -----ProHisSerLeuLeuGlyLeuTrp 355
Db 1134 CCCAGGAGTCCACACTCCCACTCCCTTCTCTGGCTGTAT 1173

RESULT 14
US-10-042-091A-1
; Sequence 1, Application US/10042091A
; Publication No. US20020142447A1
; GENERAL INFORMATION:
; APPLICANT: Darrow, Andrew
; APPLICANT: Andrade-Gordon, Patricia
; APPLICANT: Qi, Jensen
; TITLE OF INVENTION: DNA Encoding the Human Serine
; FILE REFERENCE: ORT-1031
; CURRENT APPLICATION NUMBER: US/10/042,091A
; CURRENT FILING DATE: 2002-01-08
; PRIOR APPLICATION NUMBER: US/09/387,375
; PRIOR FILING DATE: 1999-08-31
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 1613
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-042-091A-1

Alignment Scores:
Pred. No.: 2,26e-45 Length: 1613
Score: 603.50 Matches: 139
Percent Similarity: 50.00% Conservative: 38
Best Local Similarity: 39.27% Mismatches: 129
Query Match: 30.90% Indels: 48
DB: 13 Gaps: 7

US-10-037-417-46 (1-357) x US-10-042-091A-1 (1-1613)

```

```

QY 38 CysGlyArgProGluProSerAlaArgLeuValGlyGlySerAsnAlaGlnProGlyThr 57
Db 150 TGGGGCAGCCCGCATGCTCCAGTCGATGCTTGGGGCGGGATGGCGGGAGAG 209
QY 58 TrpProTrpGlnValSerLeuHisGlyGlyHisGlyGlyHisGlyGlySerLeuLeu 77
Db 210 TGGCCGTGGCAGCGAGCATCCAGCATCTCTGGGGCACAGTGTGGGGGGGTGCTCATC 269
QY 78 AlaProSerTrpValLeuSerAlaAlaHisCysPheMetThrAsnGlyThrLeuGluPro 97
Db 270 GCCCCCCAGTGGGTGTGACAGCGGGCGACTGTTTCCCGAGGAGGGCACTG----- 320
QY 98 AlaAlaGluTrpSerValLeuLeuValHisSerGlnAspGlyProLeuAspGlyAla 117
Db 321 CCAGCTGAGTACCGCGCGCTGCGTGGGGCGCTGCTGGGGCTCCACCTCGCCCCCGACG 380
QY 118 HisThrArgAlaValAlaAlaValValProAlaAsnTrpSerGlnValGluLeuGly 137
Db 381 CTCTCGTGGTGGCGGAGCGGCTGCTGCTGCCCCCGGACTCTCCGAGGAGCGGGCCCGC 440
QY 138 AlaAspLeuAlaLeuLeuArgGluAlaSerProAlaSerLeuGlyProAlaValTrpPro 157
Db 441 GCGGACTGGCACTGCTGCGAGCTGCGTGGCGCGGCTGCGCTGCGGCTCGCGTCCAAACC 500
QY 158 ValCysLeuProArgAlaSerHisArgPheValHisGlyThrAlaCysTrpAlaThrGly 177
Db 501 GTCTGCTGCGCGTGGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 560
QY 178 TrpGlyAspValGlnGluAlaAspProLeuProLeuProTrpValLeuGluValGlu 197
Db 561 TGGGGCAGCCTCCCGCCCGAGGAGTGCCTCCCGAGAGTGGCGACCGCTACAGAGGTAAAG 620

```

```

QY 198 LeuArgLeuLeuGlyGluAlaThrCysGlnCysLeuTrp-----SerGlnPro 213
Db 621 GTGCGCGTCTGCTGACTCGCGGCACCTCTACCACTGGCGGCGGACGTGCCCC 680
QY 214 GlyProPheAsnLeuThrLeuGlnLeuProGlyMetLeuCysAlaGlyTrpProGlu 233
Db 681 CAGGCTGAGCGGCAT-----GTGCTGCTGGGAGTCTGTGTGCGGCTACCCCGCAG 731
QY 234 GlyArgArgAspThrCysGlnGlyAspSerGlyGlyProLeuValCysGlnGluGlyGly 253
Db 732 GGCACACAGGAGCGCTGCCAGGCTGATCTGGGGGACCTCTGACCTGCTGCTGCTGGG 791
QY 254 ArgTrpPheGlnAlaGlyIleThrSerPheGlyPheGlyCysGlyArgArgAsnArgPro 273
Db 792 AGCTGGGTCTCTGGTGGGCGTGGTGGTGGGCAAGGGTGTGGCTGCTGCTGCTGCTGCT 851
QY 274 GlyValPheThrAlaValAlaThrTrpGluAlaTrpIleArgGluGlnValMetGlySer 293
Db 852 GGGGTCTACACAGTGTGGGCACATATAGCCCTTGGATTCAGGCTGCGCTCCTCTTAAT 911
QY 294 Glu-----ProGly----- 296
Db 912 GCTAGCGGTGAGGCTGACCTGGAGCCAGCTGCTGGGGTCCCTCAGCTCTCTGCTTCATC 971
QY 297 -----ProAlaPhe-ProThrGlnProGlnLysThrGlnSerAspCysLeuHisGlnTh 314
Db 972 CAGGCACTGCTATATACCCACATCCCTTCTGCTCGAGCCCAAGATGCTTAAAG-- 1029
QY 314 rAlaPheLeuAspSerAlaArgIleLeuLeuArgProLeuSerHisLeuSerValGlyVa 334
Db 1030 -----CTAAAGGCCACCCACCCAGCCACCCACCTTCTGGCTCTCT 1073
QY 334 lserThrGlyThrLysSerLeuValLeuProTrpLeuSer----- 347
Db 1074 CCTCTTTGGGATCACCAGCTCTGACTCCACCAACCTCTCAGGAATCTGCCATGAGT 1133
QY 348 -----ProHisSerLeuLeuGlyLeuTrp 355
Db 1134 CCCAGGAGTCCACACTCCCACTCCCTTCTCTGGCTGTAT 1173

```

```

RESULT 15
US-10-051-874-25
; Sequence 25, Application US/10051874
; Publication No. US2004000557A1
; GENERAL INFORMATION:
; APPLICANT: Padigaru, Muralidhara
; APPLICANT: Alsobrook II, John P
; APPLICANT: Colman, Steven D
; APPLICANT: Spytek, Kimberly A
; APPLICANT: Boldog, Ferenc
; APPLICANT: Vernet, Corine AM
; APPLICANT: Li, Li
; APPLICANT: Shenoy, Suresh G
; APPLICANT: Casman, Stacie J
; APPLICANT: Guo, Xiaojia Sasha
; APPLICANT: Edinger, Shlomit R
; APPLICANT: MacDougall, John R
; APPLICANT: Malyankar, Uriel M
; APPLICANT: Patturajan, Meera
; APPLICANT: Shimkets, Richard A
; APPLICANT: Pena, Carol EA
; APPLICANT: Tchernev, Velizar T
; APPLICANT: Zernusen, Bryan D
; APPLICANT: Millet, Isabelle
; APPLICANT: Miller, Charles E
; APPLICANT: Lepley, Denise M
; APPLICANT: Smithson, Glennda
; APPLICANT: Baumgartner, Jason C
; APPLICANT: Herrman, John L
; APPLICANT: Peyman, John A
; APPLICANT: Gorman, Linda
; APPLICANT: Mezes, Peter D

```

```

; APPLICANT: Kekuda, Ramesh
; APPLICANT: Taupier Jr, Raymond J
; APPLICANT: Gerlach, Valerie
; APPLICANT: Grosse, William M
; APPLICANT: Liu, Xiaohong
; APPLICANT: Ellerman, Karen
; APPLICANT: Rothenberg, Mark
; APPLICANT: Stone, David J
; APPLICANT: Burgess, Catherine E
; TITLE OF INVENTION: PROTEINS, POLYNUCLEOTIDES ENCODING THEM AND METHODS OF
; FILE REFERENCE: 21402-245
; CURRENT APPLICATION NUMBER: US/10/051,874
; CURRENT FILING DATE: 2002-09-25
; PRIOR APPLICATION NUMBER: 60/268,595
; PRIOR FILING DATE: 2001-02-14
; PRIOR APPLICATION NUMBER: 60/325,306
; PRIOR FILING DATE: 2001-09-27
; PRIOR APPLICATION NUMBER: 60/262,587
; PRIOR FILING DATE: 2001-01-18
; PRIOR APPLICATION NUMBER: 60/272,409
; PRIOR FILING DATE: 2001-02-28
; PRIOR APPLICATION NUMBER: 60/262,454
; PRIOR FILING DATE: 2001-01-18
; PRIOR APPLICATION NUMBER: 60/276,777
; PRIOR FILING DATE: 2001-03-16
; PRIOR APPLICATION NUMBER: 60/291,672
; PRIOR FILING DATE: 2001-05-17
; PRIOR APPLICATION NUMBER: 60/330,336
; PRIOR FILING DATE: 2001-10-18
; PRIOR APPLICATION NUMBER: 60/265,530
; PRIOR FILING DATE: 2001-01-31
; PRIOR APPLICATION NUMBER: 60/261,376
; PRIOR FILING DATE: 2001-01-16
; NUMBER OF SEQ ID NOS: 269
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 25
; LENGTH: 1020
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-051-874-25

```

```

Alignment Scores:
Pred. No.: 1,86e-44 Length: 1020
Score: 591.00 Matches: 125
Percent Similarity: 54.17% Conservative: 31
Best Local Similarity: 43.40% Mismatches: 102
Query Match: 30.26% Indels: 30
DB: 15 Gaps: 6

US-10-037-417-46 (1-357) x US-10-051-874-25 (1-1020)

QY 38 CysGlyArgProGluProSerAlaArgIleValGlyGlySerAsnAlaGlnProGlyThr 57
Db 172 TGGGGGAGCCCGCGCATGTCAGTCGATCGTGGGGCGCGGATGCCGGGACGGAGAG 231
QY 58 TrpProTrpGlnValSerLeuHisHisGlyGlyGlyHisIleCysGlyGlySerLeuIle 77
Db 232 TGGCCGTGGGAGCGGAGCATCCAGCATCGTGGGGGACACGCTGTCGGGGGGTGGCTATC 291
QY 78 AlaProSerTrpValLeuSerAlaAlaHisCysPheMetThrAsnGlyThrLeuGluPro 97
Db 292 GCCCCCCCGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 342
QY 98 AlaAlaGluTrpSerValLeuLeuGlyValHisSerGlnAspGlyProLeuAspGlyAla 117
Db 343 CCAGCTGAGTACCGCGTGGCGTGGCGTGGCGTGGCGTGGCGTGGCGTGGCGTGGCGTGG 402
QY 118 HisThrArgAlaValAlaAlaIleValValProAlaAsnTyrSerGlnValGluLeuGly 137
Db 403 CTCCTGGTGGCGGAGGGGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 462
QY 138 AlaAspLeuAlaLeuLeuArgLeuAlaSerProAlaSerLeuGlyProAlaValTrpPro 157

```

```

Db 463 GGGACCTGGCAGTCGTCAGTCGCGCGGTCGCTGCGCTGAGCGTCGCGTCACACCC 522
QY 158 ValCysLeuProArgAlaSerHisArgPheValHisGlyThrAlaCysTrpAlaThrGly 177
Db 523 GTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 582
QY 178 TrpGlyAspValGlnGluAlaAspProLeuProLeuProTrpValLeuGlnGluValGlu 197
Db 583 TGGGGCAGCCTCCCGCCAGGAGTCCCTCCAGAGTGGCGACCGCTACAAAGGAGTAAGG 642
QY 198 LeuArgLeuLeuGlyGluAlaThrCysGlnCysLeuTrp-----SerGlnPro 213
Db 643 GTGCGCTGCTGGAGTCGCGCAGCTGCGCAGCTGCGCAGCTGCGCAGCTGCGCAGCTGCG 702
QY 214 GlyProPheAsnLeuThrLeuGlnIleLeuProGlyMetLeuCysAlaGlyTyrProGlu 233
Db 703 CAGGCTGAGCGCATT-----GTGCTGCTGGGAGTCTGTGTGCGGCTACCCCGAG 753
QY 234 GlyArgArgAspThrCysGlnGlyAspSerGlyGlyProLeuValCysGluGluGlyGly 253
Db 754 GGGCACAAGGACGCTGCGAGGGTGAATCTGGGGGACCTCTGACCTGCTGCGTGGGCTGGG 813
QY 254 ArgTrpPheGlnAlaGlyIleThrSerPheGlyPheGlyCysGlyArgArgAsnArgPro 273
Db 814 AGCTGGGTCCTGGTGGGCGTGGTGGGCGAGGGTGTGTGCTGCTGCGCTGCGCAGCTGCG 873
QY 274 GlyValPheThrAlaValAlaThrTyrGluAlaTrpIleArg----- 287
Db 874 GGGGTCTACACCAAGTGTGGCCACATATAGCCCTGGATTGAGTTCAGGCTGCGCTGCGT 933
QY 288 -----GluGlnValMetGlySer-----Glu 294
Db 934 TGCTAGCCGCTGAGGCTGAGTGGAGCCAGCTGCTGGGGTCCCTCAGCCTCCTGGTTCAT 993
QY 295 ProGlyProAla---PheProThr 301
Db 994 CCAGGACCTGCTATACCCACACA 1017

```

Search completed: February 25, 2004, 22:14:22
Job time: 414 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: February 25, 2004, 18:03:06 ; Search time 3777 Seconds

(without alignments)
4096.760 Million cell updates/sec

Title: US-10-037-417-46

Perfect score: 357

Sequence: 1 MAQKGVLPGLGAVANSDS.....TKSLVLPWLSPHSLGLMGF 357

Scoring table: OLIGO
Xgapop 60.0 , Xgapext 60.0
Ygapop 60.0 , Ygapext 60.0
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 3470272 seqs, 21671516995 residues

Word size: 1

Total number of hits satisfying chosen parameters: 6934743

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlp
-Q=/cgn2_1/USPTO_spool_p/US10037417/runat_24022004_101108_18083/app_query_fasta_1.519
-DB=GenEmbl -QFMT=fastap -SUFFIX=olip2n.rge -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=oligo -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR SCORE=quality -THR MIN=1 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptp
-NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10037417@cgn_1_4958@runat_24022004_101108_18083 -NCPU=6 -ICPU=3
-NO MMAP -LARGQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database :

GenEmbl.*

1: gb.ba.*
2: gb.htg.*
3: gb.in.*
4: gb.om.*
5: gb.ov.*
6: gb.pat.*
7: gb.ph.*
8: gb.pl.*
9: gb.pr.*
10: gb.ro.*
11: gb.sts.*
12: gb.sy.*
13: gb.un.*
14: gb.vi.*
15: em.ba.*
16: em.fun.*
17: em.hum.*
18: em.in.*
19: em.mu.*
20: em.om.*
21: em.or.*
22: em.ov.*
23: em.pat.*
24: em.ph.*
25: em.pl.*
26: em.ro.*
27: em.sts.*
28: em.un.*

29: em.vi.*
30: em.htg.hum.*
31: em.htg.inv.*
32: em.htg.other.*
33: em.htg.mus.*
34: em.htg.pln.*
35: em.htg.rod.*
36: em.htg.mam.*
37: em.htg.vrt.*
38: em.sy.*
39: em.htg.hum.*
40: em.htg.mus.*
41: em.htg.other.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query			DB	ID	Description
		Match	Length	%			
1	357	100.0	1102	6	AX512287	AX512287	Sequence
2	357	100.0	1102	6	AX512289	AX512289	Sequence
3	272	76.2	2457	6	AX360096	AX360096	Sequence
4	272	76.2	2681	6	AX480935	AX480935	Sequence
5	261	73.1	2810	6	BD127529	BD127529	Primer fo
6	261	73.1	2810	9	AK075142	AK075142	Homo sapi
7	178	49.9	537	6	AX342934	AX342934	Sequence
8	132	37.0	670	6	BD125219	BD125219	Primer fo
9	132	37.0	670	6	BD126485	BD126485	Primer fo
10	128	35.9	456	6	AX342936	AX342936	Sequence
11	93	26.1	127769	9	AC009088	AC009088	Homo sapi
12	93	26.1	195476	2	AC135044	AC135044	Homo sapi
13	35	9.8	232119	2	AC106629	AC106629	Rattus no
14	35	9.8	244161	2	AC117170	AC117170	Rattus no
15	29	8.1	168064	9	AC093520	AC093520	Homo sapi
16	23	6.4	64328	2	AC124816	AC124816	Mus muscu
17	23	6.4	185788	10	AC124461	AC124461	Mus muscu
18	23	6.4	185559	2	AC124602	AC124602	Mus muscu
19	23	6.4	219763	10	AC093175	AC093175	Mus muscu
20	19	5.3	71177	2	AC101465	AC101465	Mus muscu
21	17	4.8	6207	9	BC035589	BC035589	Homo sapi
22	17	4.8	7604	9	AB002294	AB002294	Human mRN
23	17	4.8	182230	9	AC135050	AC135050	Homo sapi
24	16	4.5	1161	6	AX675581	AX675581	Sequence
25	16	4.5	1726	6	AX675579	AX675579	Sequence
26	16	4.5	1796	6	AX098193	AX098193	Sequence
27	16	4.5	1800	9	BC001462	BC001462	Homo sapi
28	16	4.5	1834	6	AX335777	AX335777	Sequence
29	16	4.5	1834	6	AX336076	AX336076	Sequence
30	16	4.5	1834	6	AX336340	AX336340	Sequence
31	16	4.5	1834	6	AX409567	AX409567	Sequence
32	16	4.5	1834	6	AX474697	AX474697	Sequence
33	16	4.5	1834	9	HUMPR0S	HUMPR0S	Homo sapien
34	16	4.5	1835	6	AX098215	AX098215	Sequence
35	16	4.5	7008	9	HSU33446	HSU33446	Human prost
36	16	4.5	71177	2	AC101465	AC101465	Mus muscu
37	16	4.5	195476	2	AC135044	AC135044	Homo sapi
38	14	3.9	1931	10	MMX99571	MMX99571	M.musculus
39	14	3.9	1963	10	BC037085	BC037085	Mus muscu
40	14	3.9	1968	10	BC057921	BC057921	Mus muscu
41	14	3.9	1980	10	BC049867	BC049867	Mus muscu
42	13	3.6	506	6	AR263894	AR263894	Sequence
43	13	3.6	603	10	AB074296	AB074296	Mus muscu
44	13	3.6	711	6	AX763030	AX763030	Sequence
45	13	3.6	792	6	AR337567	AR337567	Sequence

ALIGNMENTS

```
AX512287
LOCUS AX512287 1102 bp DNA linear PAT 27-SEP-2002
DEFINITION Sequence 43 from Patent WO02053742.
ACCESSION AX512287
VERSION AX512287.1 GI:23392631
KEYWORDS Homo sapiens (human)
SOURCE Homo sapiens
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1
AUTHORS Kekuda,R., Alsobrook,J.P., Tchiernev,V.T., Liu,X., Spytek,K.A.,
Patturajan,M., Grosse,W.M., Lepley,D.M., Burgess,C.E., Vernet,C.A.,
Li,L., Gorman,L., Edinger,S., Sciore,P., Ellerman,K., Malyankar,U.,
Rothenberg,M., Stone,D., Boldog,F., Shenoy,S. and Anderson,D.
Proteins and nucleic acids encoding same
PATENT: WO 02053742-A 43 11-JUL-2002;
Curagen Corporation (US)
FEATURES
source
Location/Qualifiers
1..1102
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"
ORIGIN
Alignment Scores:
Pred. No.: 0 Length: 1102
Score: 357.00 Matches: 357
Percent Similarity: 100.00% Conservatives: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 6 Gaps: 0
US-10-037-417-46 (1-357) x AX512287 (1-1102)
QY 1 MetAlaGlnLysGlyValLeuGlyProGlyGlnLeuGlyAlaValAlaAAsnSerAspSer 20
DB 19 ATGGCCAGAGGGGTCTCTGGGCGCTGGGAGCTGGGGGCTGTGGCCAAATTCGACTCA 78
QY 21 TyrSerLeuTyrGlyLeuValProSerGlyProAlaArgGlyProProTyrCysGlyArg 40
DB 79 TACTCACTTTACGGGTGTGGTGGCTGGAGCCGGCTAGGGGGCCCCCGTACTGCGGGGCGC 138
QY 41 ProGluProSerAlaAlaGlnValGlyGlySerAsnAlaGlnProGlyThrTrpProTrp 60
DB 139 CCTGAGCCCTCGGCCCGCATCGTGGGGGCTCAACGCGAGCGGGGACCTGGCCTTGG 198
QY 61 GlnValSerLeuHisGlyGlyGlyHisIleCysGlyGlySerLeuLeuAlaProSer 80
DB 199 CAAGTGAGCCTGCACCATGGAGGTGGCCACATCTGCGGGGGCTCCCTCATCGCCCTCC 258
QY 81 TrpValLeuSerAlaAlaHisCysPheMetThrAsnGlyThrLeuGluProAlaAlaGlu 100
DB 259 TGGGTCTCTCCGCTGCTCACTGTTTCATGACGATGGGAGCTTGGAGCCGCGGGCGAG 318
QY 101 TrpSerValLeuLeuGlyValHisSerGlnAspGlyProLeuAspGlyAlaHisThrArg 120
DB 319 TGGTGGTACTGTGGGGCGTGCACTCCAGGACGGGGCCCCCTGGAGCGGGCGCACCCCGC 378
QY 121 AlaValAlaAlaValValProAlaAsnTyrSerGlnValGluLeuGlyAlaAspLeu 140
DB 379 GCAGTGGCGGCATCGTGGTCCCGCCCAACTACAGCAAGTGGAGCTGGGCGCGGACCTG 438
QY 141 AlaLeuLeuArgLeuAlaSerProAlaSerLeuGlyProAlaValTrpProValCysLeu 160
DB 439 GCCCTGTGCGCCTGGCCCTCACCCGACAGCTGGGGCCCCCGCGCTGTGGCTGTGCTG 498
QY 161 ProArgAlaSerHisArgPheValHisGlyThrAlaCysTrpAlaThrGlyTrpGlyAsp 180
DB 499 CCGCGGCGCTCACCGCTTGTGTCACGGCACCGCCCTGTGGGCCACCGCGTGGGGAGAC 558
QY 181 ValGlnGluAlaAspProLeuProLeuProTrpValLeuGlnGluValGluLeuArgLeu 200
```

```
DB 559 GTCCAGGAGGAGATCCTCTGCCTCTCCCTCGGTGCTCTACAGAACTGAGCTAAGGCTG 618
QY 201 LeuGlyGluAlaThrCysGlnCysLeuTyrSerGlnProGlyProPheAsnLeuThrLeu 220
DB 619 CTGGGCGAGGCCACCTGTCAATGCTCTTACAGCAGCCCGGTCCCTTCACTCACTCTC 678
QY 221 GlnIleLeuProGlyMetLeuCysAlaGlyTyrProGluClyArgAspThrCysGln 240
DB 679 CAGATATTGCCAGGATGCTGTGTGTGCTACCCAGAGGCCCGCAGGACACCTGCCAG 738
QY 241 GlyAspSerGlyGlyProLeuValCysGluGluGlyGlyArgTrpPheGlnAlaGlyIle 260
DB 739 GGTGACTCTGGGGGGCCCCCTGGTCTGTGAGAAAGCGGCGCTGTTCCAGGACGAATC 798
QY 261 ThrSerPheGlyPheGlyCysGlyArgArgAsnArgProGlyValPheThrAlaValAla 280
DB 799 ACCAGCTTTGGTTGGCTGTGGACGAGAAACCGCCCTGGAGTTTTCACCTGCTGTGGCT 858
QY 281 ThrTyrGluAlaTrpIleArgGluGlnValMetGlySerGluProGlyProAlaPhePro 300
DB 859 ACCTATGAGCATGGATACGGGAGCAGGTATGGTTTCAGAGCTTGGGCTGCTTTTCCC 918
QY 301 ThrGlnProGlnLysThrGlnSerAspCysLeuHisGlnThrAlaPheLeuAspSerAla 320
DB 919 ACCAGCCCCCAGAGACCCAGTCAGATTGTTTACATCAACCGCATTCCTGGATTCTGCC 978
QY 321 ArgIleLeuLeuArgProLeuSerHisIleSerValGlyValSerThrGlyThrIlySer 340
DB 979 AGAATCTCTTTGAGGCGCTTGTCCCATATATCATAGTAGAGTCTCACTGGGACCAAAAGC 1038
QY 341 LeuValLeuProTrpLeuSerProHisSerLeuLeuGlyLeuTrpGlyPhe 357
DB 1039 CTTGTCTCTCTCTGGCTCTCTCCACACTCTCTCTCTGGGCTCTGGGGGTTT 1089
RESULT 2
AX512289
LOCUS AX512289 1102 bp DNA linear PAT 27-SEP-2002
DEFINITION Sequence 45 from Patent WO02053742.
ACCESSION AX512289
VERSION AX512289.1 GI:23392632
KEYWORDS Homo sapiens (human)
SOURCE Homo sapiens
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1
AUTHORS Kekuda,R., Alsobrook,J.P., Tchiernev,V.T., Liu,X., Spytek,K.A.,
Patturajan,M., Grosse,W.M., Lepley,D.M., Burgess,C.E., Vernet,C.A.,
Li,L., Gorman,L., Edinger,S., Sciore,P., Ellerman,K., Malyankar,U.,
Rothenberg,M., Stone,D., Boldog,F., Shenoy,S. and Anderson,D.
Proteins and nucleic acids encoding same
PATENT: WO 02053742-A 45 11-JUL-2002;
Curagen Corporation (US)
FEATURES
source
Location/Qualifiers
1..1102
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"
ORIGIN
Alignment Scores:
Pred. No.: 0 Length: 1102
Score: 357.00 Matches: 357
Percent Similarity: 100.00% Conservatives: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 6 Gaps: 0
US-10-037-417-46 (1-357) x AX512289 (1-1102)
QY 1 MetAlaGlnLysGlyValLeuGlyProGlyGlnLeuGlyAlaValAlaAAsnSerAspSer 20
DB 19 ATGGCCAGAGGGGTCTCTGGGCGCTGGGAGCTGGGGGCTGTGGCCAAATTCGACTCA 78
QY 21 TyrSerLeuTyrGlyLeuValProSerGlyProAlaArgGlyProProTyrCysGlyArg 40
DB 79 TACTCACTTTACGGGTGTGGTGGCTGGAGCCGGCTAGGGGGCCCCCGTACTGCGGGGCGC 138
QY 41 ProGluProSerAlaAlaGlnValGlyGlySerAsnAlaGlnProGlyThrTrpProTrp 60
DB 139 CCTGAGCCCTCGGCCCGCATCGTGGGGGCTCAACGCGAGCGGGGACCTGGCCTTGG 198
QY 61 GlnValSerLeuHisGlyGlyGlyHisIleCysGlyGlySerLeuLeuAlaProSer 80
DB 199 CAAGTGAGCCTGCACCATGGAGGTGGCCACATCTGCGGGGGCTCCCTCATCGCCCTCC 258
QY 81 TrpValLeuSerAlaAlaHisCysPheMetThrAsnGlyThrLeuGluProAlaAlaGlu 100
DB 259 TGGGTCTCTCCGCTGCTCACTGTTTCATGACGATGGGAGCTTGGAGCCGCGGGCGAG 318
QY 101 TrpSerValLeuLeuGlyValHisSerGlnAspGlyProLeuAspGlyAlaHisThrArg 120
DB 319 TGGTGGTACTGTGGGGCGTGCACTCCAGGACGGGGCCCCCTGGAGCGGGCGCACCCCGC 378
QY 121 AlaValAlaAlaValValProAlaAsnTyrSerGlnValGluLeuGlyAlaAspLeu 140
DB 379 GCAGTGGCGGCATCGTGGTCCCGCCCAACTACAGCAAGTGGAGCTGGGCGCGGACCTG 438
QY 141 AlaLeuLeuArgLeuAlaSerProAlaSerLeuGlyProAlaValTrpProValCysLeu 160
DB 439 GCCCTGTGCGCCTGGCCCTCACCCGACAGCTGGGGCCCCCGCGCTGTGGCTGTGCTG 498
QY 161 ProArgAlaSerHisArgPheValHisGlyThrAlaCysTrpAlaThrGlyTrpGlyAsp 180
DB 499 CCGCGGCGCTCACCGCTTGTGTCACGGCACCGCCCTGTGGGCCACCGCGTGGGGAGAC 558
QY 181 ValGlnGluAlaAspProLeuProLeuProTrpValLeuGlnGluValGluLeuArgLeu 200
```

```

QY      21  TyrSerLeuTyrGlyLeuValProSerGlyProAlaArgGlyProProTyrCysGlyArg 40
DB      79  TACTACACITTTACGGTTGGTGGCTCGGTCGACCGCTAGGGGGCCCCCGTACTGGGGGCG 138
QY      41  ProGluProSerAlaArgIleValGlyGlySerAsnAlaGlnProGlyThrTrpProTrp 60
DB      139  CTGAGCCCTCGGCCCGCATCGTGGGGGGCTCAACGCGCAGCGCGGACCTGGCCCTGG 198
QY      61  GlnValSerLeuHisGlyGlyGlyHisIleCysGlyGlySerLeuIleAlaProSer 80
DB      199  CAAGTGAGCTGCACCATGAGGTGGCCCATCTGCGGGGCTCCCTCATCGGCCCTCC 258
QY      81  TrpValLeuSerAlaAlaHisCysPheMetThrAsnGlyThrLeuGluProAlaAlaGlu 100
DB      259  TGGGTCCTCTCCGCTGCTCACTGTTTCATGACGANTGGGAGCTGTGAGCCCGGCGGAG 318
QY      101  TrpSerValLeuLeuGlyValHisSerGlnAspGlyProLeuAspGlyAlaHisThrArg 120
DB      319  TGGTCGTACTGCTGGCGGTGCACTCCAGAGCGGGCCCTGGACGCGCGCACACCCGC 378
QY      121  AlaValAlaAlaIleValValProAlaAsnTyrSerGlnValGluLeuGlyAlaAspLeu 140
DB      379  GCAGTGGCCGCCATCGGTGGTGGCGGCCAATACAGCCAGTGAGCTGGCGGCCGACCTG 438
QY      141  AlaLeuLeuArgLeuAlaSerProAlaSerLeuGlyProAlaValTrpProValCysLeu 160
DB      439  GCCCTGCTGGCCCTGGCTCACCGCAGCTGGGCCCGCCGCTGGCTGTGCTGCTGCTG 498
QY      161  ProArgAlaSerHisArgPheValHisGlyThrAlaCysTrpAlaThrGlyTrpGlyAsp 180
DB      499  CCGCGCCCTCACACCGCTCTGTGACGGCACCGCCCTGCTGGGCCACCGCTGGGGAGAC 558
QY      181  ValGlnGluAlaAspProLeuProTrpValLeuGlnValGluLeuAlaTrpLeu 200
DB      559  GTCCAGAGGAGATCTCTGCTCTCCCTGGGTGGTACAGGAAGTGGAGCTAAGGCTG 618
QY      201  LeuGlyGluAlaThrCysGlnCysLeuTyrSerGlnProGlyProPheAsnLeuThrLeu 220
DB      619  CTGGGCGAGGCCACTGTCAATGTCTCTACAGCCAGCCCGCTCCCTTCACTCACTCTC 678
QY      221  GlnIleLeuProGlyMetLeuCysAlaGlyTyrProGluGlyArgArgAspThrCysGln 240
DB      679  CAGATAATTGCCAGGATGCTGTGTGCTGGCTACCCAGAGGGCCCGCAGGGACACCTGCCAG 738
QY      241  GlyAspSerGlyGlyProLeuValCysGluGluGlyArgTrpPheGlnAlaGlyIle 260
DB      739  GGTCACTCTGGGGGGCCCTGCTGTGAGGAAGCGGCCCTGCTTCCAGCAGGAATC 798
QY      261  ThrSerPheGlyPheGlyCysGlyArgArgAsnArgProGlyValPheThrAlaValAla 280
DB      799  ACCAGCTTTGGTTGGCTGTGGAGCGGAGAAACCGCCCTGGAGTTTCACTGCTGTGGCT 858
QY      281  ThrTyrGluAlaIleAlaGluGlnValMetGlySerGluProGlyProAlaPhePro 300
DB      859  ACCTATGAGGATGATATGAGGAGCAGGTGATGGTTTCAGAGCCCTGGCCCTTCCCTCC 918
QY      301  ThrGlnProGlnLysThrGlnSerAspCysLeuHisGlnThrAlaPheLeuAspSerAla 320
DB      919  ACCAGCCCCAGAGACCCAGTCAGATTGTTTACATCAACCGCATCTCTGGATTCTGCC 978
QY      321  ArgIleLeuLeuArgProLeuSerHisIleSerValGlyValSerThrGlyThrIysSer 340
DB      979  AGAATCCCTTTGAGGCCCTTGTCCCATATATCAGTAGGAGTCTCACTGGGACCAAAAGC 1038
QY      341  LeuValLeuProTrpLeuSerProHisSerLeuLeuGlyLeuTrpGlyPhe 357
DB      1039  CTTGTCTCTCCCTGGCTCTCTCCACACTCTCTCTCTGGGCCCTCTGGGGGTTTC 1089

```

RESULT 3

AX360096

LOCUS

DEFINITION Sequence 52 from Patent WO0200860.

2457 bp

DNA

linear

PAT 13-FEB-2002

```

ACCESSION AX360096
VERSION AX360096.1 GI:18675722
KEYWORDS Homo sapiens (human)
SOURCE Homo sapiens
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1
AUTHORS Plowman,G., Whyte,D., Sudarsanam,S., Manning,G., Caenepeel,S. and
Charydzak,G.
TITLE Novel proteases
JOURNAL Patent: WO 0200860-A 52 03-JAN-2002;
FEATURES
source
1..2457
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"
ORIGIN
Alignment Scores:
Pred. No.: 2,41e-265 Length: 2457
Score: 272.00 Matches: 272
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 76.19% Indels: 0
DB: Gaps: 6
US-10-037-417-46 (1-357) x AX360096 (1-2457)
QY 38 CysGlyArgProGluProSerAlaArgIleValGlyGlySerAsnAlaGlnProGlyThr 57
DB 112 TGGGGGCGCCCTGAGCCCTCGGCCCGCATCTGGGGGGGCTCAAAACGCGCAGCGGGGACC 171
QY 58 TrpProTrpGlnValSerLeuHisGlyGlyGlyHisIleCysGlyGlySerLeuIle 77
DB 172 TGGCCCTGGCAAGTAGAGCTGCACCATGGAGGTGGCCACATCTGGCGGGGCTCCCTCATC 231
QY 78 AlaProSerTrpValLeuSerAlaAlaHisCysPheMetThrAsnGlyThrLeuGluPro 97
DB 232 GCCCCCTCTCGGTCTCTCCGCTGCTCACTGTTTCATGACGAATGGGACGCTGGAGCCC 291
QY 98 AlaAlaGluTrpSerValLeuLeuGlyValHisSerGlnAspGlyProLeuAspGlyAla 117
DB 292 GCGGCGGAGTGGTGGGTGCTGCTGGGGCGTGCATCCCGAGACGGGCCCTTGGACGGGCG 351
QY 118 HisThrArgAlaValAlaAlaIleValValProAlaAsnTyrSerGlnValGluLeuGly 137
DB 352 CACACCGCGCAGTGGCGCCCATCTGCTGTCGCGCCCAACTACAGCCCAAGTGGAGCTGGGC 411
QY 138 AlaAspLeuAlaLeuLeuArgLeuAlaSerProAlaSerLeuGlyProAlaValTrpPro 157
DB 412 GCCAGCTGGCCCTGCTGCGCTGCGCTCAACCGCTTGTGACCGCACCGCTGCTGGCCACCGGC 471
QY 158 ValCysLeuProArgAlaSerHisArgPheValHisGlyThrAlaCysTrpAlaThrGly 177
DB 472 GTCTGCTTGGCCCCCGCTCACACCGCTTGTGACCGCACCGCTGCTGGCCACCGGC 531
QY 178 TrpGlyAspValGlnGluAlaAspProLeuProLeuProTrpValLeuGlnValGlu 197
DB 532 TGGGGAGACGCTCCAGGAGGACATCTCTGCTCTCCCTGGGTGTCTACAGGAAGTGGAG 591
QY 198 LeuArgIleLeuGlyGluAlaThrCysGlnCysLeuTyrSerGlnProGlyProPheAsn 217
DB 592 CTAAGGCTGCTGGGCGAGGCCACCTGTCAATGTCTCTACAGCCACCGCGGCTCCCTTTCAAC 651
QY 218 LeuThrLeuGlnIleLeuProGlyMetLeuCysAlaGlyTyrProGluGlyArgArgAsp 237
DB 652 CTCACCTTCCAGATATTGCCAGGGATGCTGTGCTGTGCTACCCAGAGGGCGGCGAGGAC 711
QY 238 ThrCysGlnGlyAspSerGlyProLeuValCysGluGlyGlyArgTrpPheGln 257
DB 712 ACCTGCCAGGCTGACTCTGGGGGGCCCCCTGCTGTGTGAGGAAGCGCGCGCTGGTTCCAG 771

```

QY 258 AlaGlyThrSerPheGlyPheGlyCysGlyArgArgSerProGlyValPheThr 277
 Db 772 GCAGGAATCACCAGCTTTGGCTTTGGCTGTGGACGGAGAACCCGCTGGAGTTTTCAC 831

QY 278 AlaValAlaThrTyrGluAlaTyrPheGlyValMetGlySerGluProGlyPro 297
 Db 832 GCTGTGGCTACCTATGAGGAGTATGATACGGAGAGAGTGTGGTTTCAGAGCTGGGCT 891

QY 298 AlaPheProThrGlnProGlnLysThrGlnSerAsp 309
 Db 892 GCCTTTCCACCCAGCCAGAGACCCAGTGCAGAT 927

RESULT 4
 LOCUS AX480935 2681 bp DNA linear PAT 12-AUG-2002
 DEFINITION Sequence 31 from Patent WO0246383.
 ACCESSION AX480935
 VERSION AX480935.1 GI:22217574
 KEYWORDS Homo sapiens (human)
 SOURCE
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1
 AUTHORS Yue, H., Aizumi, Y., Kallick, D.A., Baughn, M.R., Griffin, J.A., Swarnakar, A., Lai, P.G., Walla, N.K., Rafaila, A.J., Gandhi, A.R., Au-Yang, J., Elliott, V.S., Ramkumar, J., Thangavelu, K., Lu, Y., Warren, B.A., Lu, D.A., Lee, E.A., Tribouley, C.M., Arvizu, C., Deleane, A.M., Yao, M.G., Khan, F.A. and Sanjanwala, M.M.
 TITLE Protein modification and maintenance molecules
 JOURNAL Patent: WO 0246383-A 31 JUN-2002;
 Incyte Genomics, Inc. (US)
 FEATURES
 Location/Qualifiers
 1..2681
 /organism="Homo sapiens"
 /mol_type="unassigned DNA"
 /db_xref="taxon:9606"
 /note="Incyte ID No: 2751509CB1"

ORIGIN
 Alignment Scores: 2,57e-265 Length: 2681
 Pred. No.: 272.00 Matches: 272
 Score: 100.00% Conservative: 0
 Percent Similarity: 100.00% Mismatches: 0
 Best Local Similarity: 76.19% Indels: 0
 Query Match: 6 Gaps: 0
 DB:

US-10-037-417-46 (1-357) x AX480935 (1-2681)

QY 38 CysGlyArgProGluProSerAlaArgTyrGlySerAsnAlaGlnProGlyThr 57
 Db 112 TGGCGGGCGCCCTGAGCCCTCGCCGCGATCGTGGGGGCTCAACCGCGAGCGGGGACC 171

QY 58 TrpProThrGlnValSerLeuHisGlyGlyGlyHisIleCysGlyGlySerLeu 77
 Db 172 TGGCGCTTGGCAAGTGGCTGACCATGGAGTGGCCACATCTGGCGGGGCTCCCTCATC 231

QY 78 AlaProSerTrpValLeuSerAlaAlaHisCysPheMetThrAsnGlyThrLeuGluPro 97
 Db 232 GCCCCCTCTGGGTCCTCTCCGCTGCTCACTGTTTCATGCAATGGAGCGCTGGAGGCC 291

QY 98 AlaAlaGluTrpSerValLeuGlyValHisSerGlnAspGlyProLeuAspGlyAla 117
 Db 292 GCGGCGGAGTGGTGGTACTCTGGGGGTGACCTCCAGGAGCGGCGGCTGGAGCGGCG 351

QY 118 HisThrArgAlaValAlaIleValValProAlaSerTyrSerGlnValGluLeuGly 137
 Db 352 CACACCCCGCAGTGGCGCCATCTGTGGTGGCGCCACTACAGCCAGTGGAGCTGGC 411

QY 138 AlaAspLeuAlaLeuLeuArgLeuAlaSerProAlaSerLeuGlyProAlaValTrpPro 157

Db 412 GCCGACCTGGCCCTGCTGGCCCTGGCCCTCACCCGCCAGCGTGGGCCCGCGCTGTGGCT 471

QY 158 ValCysLeuProArgAlaSerHisArgPheValHisGlyThrAlaCysTrpAlaThrGly 177
 Db 472 GTCTGCTGCCCGCGCTTCTGTCACACCGCTTCTGTCACGCGACCGCTGTGGGCCACGGC 531

QY 178 TrpGlyAspValGlnGluAlaAspProLeuProLeuProTrpValLeuGlnGluValGlu 197
 Db 532 TGGGAGACGCTCCAGGAGGAGATCTCTGCTCTCCCTTGGGTGTACAGGAAGTGGAG 591

QY 198 LeuArgLeuLeuGlyGluAlaThrCysGlnCysLeuTyrSerGlnProGlyProPheAsn 217
 Db 592 CTAAGGCTGTGGCGGAGGAGCCACTGTCAATGTCTTACAGCCAGCCCGCTCCCTTCAAC 651

QY 218 LeuThrLeuGlnIleLeuProGlyMetLeuCysAlaGlyTyrProGluGlyArgArgAsp 237
 Db 652 CTCACTCTCCAGATATTGCCAGGATGCTGTGTGCTTACCCAGAGGCGCGAGGAG 711

QY 238 ThrCysGlnGlyAspSerGlyGlyProLeuValCysGluGluGlyGlyArgTrpPheGln 257
 Db 712 ACCTGCCAGGCTGACTCTCTGGGGGCGCTGTGTGTGAGGAAGCGCGCTGTGTTCCAG 771

QY 258 AlaGlyThrSerPheGlyPheGlyCysGlyArgArgAsnArgProGlyValPheThr 277
 Db 772 GCAGGAATCACCAGCTTTGGCTTTGGCTGTGGACGGAGAACCCGCTGGAGTTTTCAC 831

QY 278 AlaValAlaThrTyrGluAlaTyrPheGlyValMetGlySerGluProGlyPro 297
 Db 832 GCTGTGGCTACTATGAGGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 891

QY 298 AlaPheProThrGlnProGlnLysThrGlnSerAsp 309
 Db 892 GCCTTTCCACCCAGCCAGAGACCCAGTGCAGAT 927

RESULT 5
 LOCUS BD127529 2810 bp DNA linear PAT 18-SEP-2002
 DEFINITION Primer for synthesizing full-length cDNA and use thereof.
 ACCESSION BD127529
 VERSION BD127529.1 GI:23222474
 KEYWORDS JP 2002017375-A/2960.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (bases 1 to 2810)
 AUTHORS Ota, T., Nishikawa, T., Isogai, T., Hayashi, K., Ishii, S., Kawai, Y., Wakamatsu, A., Sugiyama, T., Nagai, K., Kojima, S., Otsuki, T. and Koga, H.
 TITLE Primer for synthesizing full-length cDNA and use thereof
 JOURNAL Patent: JP 2002017375-A 2960 22-JAN-2002;
 HELIX RESEARCH INSTITUTE
 COMMENT OS Homo sapiens (human)
 PN JP 2002017375-A/2960
 PD 22-JAN-2002
 PF 07-JUN-2000 JP 20020253172
 PI TOSHIO OTA, TETSUO NISHIKAWA, TAKAO ISOGAI, KOJI HAYASHI, SHIZUKO ISHII,
 PI YURI KAWAI, AI WAKAMATSU, TOMOYASU SUGIYAMA, KEIICHI NAGAI, PI SHINICHI KOJIMA,
 PI TETSUJI OTSUKI, HISASHI KOGA
 PC C12N15/09, C07K14/47, C07K16/18, C12N1/15, C12N1/19, C12N1/21, C12N5/ PC 10,
 C12P21/02, C12Q1/68//C12P21/08, G06F17/30, C12N15/00, C12N5/00 CC
 Primer for synthesizing full-length cDNA and use thereof FH Key
 Location/Qualifiers
 FT CDS (300)...(2597).
 1..2810
 /organism="Homo sapiens"
 /mol_type="genomic DNA"
 /db_xref="taxon:9606"

[illegible]

QY 69 GlyHisIleCysGlySerLeuIleAlaProSerTrpValLeuSerAlaAlaHisCys 88
 Db 237 GGCCACATCTGGGGGCTCCCTCATCGCCCTCTGGGTCTCTCCGCTGCTACTGT 296
 QY 89 PheMetThrAsnGlyThrLeuGluProAlaAlaGluTrpSerValLeuLeuGlyValHis 108
 Db 297 TTCATGACGAATGGAGCGCTGGAGCCCGCGCGAGTGGTTCGTACTGCTGGCGGTGCAC 356
 QY 109 SerGlnAspGlyProLeuAspGlyAlaHisThrArgAlaValAlaAlaIleValPro 128
 Db 357 TCCAGAGAGCGGCCCTGGAGCGGGCGCACACCGCGCAGTGGCGGCATCGTGGTGGCG 416
 QY 129 AlaAsnTrpSerGlnValGluLeuGlyAlaAspLeuAlaLeuArgLeuAlaSerPro 148
 Db 417 GCCAACTACAGCCAACTGGAGCTGGGCGCCGACCTGGCCCTGCTGGCCCTGACCC 476
 QY 149 AlaSerLeuGlyProAlaValTrpProValCysLeuProArgAlaSerHisArgPheVal 168
 Db 477 GCCAGCTGGGCGCCCGCGTGTGGCTGTCTGGCTGGCCCGCGCTCACACCGCTTGGTG 536
 QY 169 HisGlyThrAlaCysTrpAlaThrGlyTrpGlyAspValGlnGluAlaAspProLeuPro 188
 Db 537 CACGGCACCGCTGTGGGCCACCGGCTGGGAGACGTCAGGAGGCGAGATCTCTGGCT 596
 QY 189 LeuProTrpValLeuGlnGluValGluLeuArgLeuLeuGlyGluAlaThrCysGlnCys 208
 Db 597 CTCCCTGGGTCTACAGGAATGGAGTAAAGCTGGTGGGCGAGCCACCTGTCAATGT 656
 QY 209 LeuTrpSerGlnProGlyProPheAsnLeuThrLeuGlnIleLeuProGlyMetLeuCys 228
 Db 657 CFTCAAGCAGCCCGGCTCCCTCAACCTCACTCCAGATATTCAGGAGGATGTGTGT 716
 QY 229 AlaGlyTrpProGluGlyArgAspThrCysGlnGlyAspSerGlyClyProLeuVal 248
 Db 717 GCTGGCTACCCAGAGCGCGAGGACACCTGCGAGGCGACTCTGGGGGGCCCTGGTGC 776
 QY 249 CysGluGluGlyClyArgTrpPheGlnAlaGlyIleThrSerPheGlyPheGlyCysGly 268
 Db 777 TGTGAGGAAGCGCGCGCTGGTTCCAGGCGAGGAATCACAGCTTGGCTTGGCTTGGGA 836
 QY 269 ArgArgAsnArgProGlyValPheThrAlaValAlaAlaThrTrpGluAlaTrpIleArgGlu 288
 Db 837 CGAGAAACCGCCCTGGAGTTTCACTGCTGTGGCTACCTATGAGGCATGGATACGGAG 896
 QY 289 GlnValMetGlySerGluProGlyProAlaPheProThrGlnProGlnLysThrGlnSer 308
 Db 897 CAGGTGATGGTTTCAGAGCTGGGCTGCTTTCCACCCAGCCCGCCAGAGACCCAGTCA 956
 QY 309 Asp 309
 Db 957 GAT 959

RESULT 7
 AX342934/c
 LOCUS
 DEFINITION
 ACCESSION
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM
 Homo sapiens
 Homo sapiens
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1
 XIAO, Y. and MOROZOV, V.
 Regulation of human prostatic-like serine protease
 Patent: WO 0198467-A 1 27-DEC-2001;
 Bayer Aktiengesellschaft (DE)
 Location/Qualifiers
 1..537
 /organism="Homo sapiens"
 /mol_type="unassigned DNA"
 /db_xref="taxon:9606"
 source

ORIGIN
 Alignment Scores:
 Pred. No.: 1.59e-170 Length: 537
 Score: 178.00 Matches: 178
 Percent Similarity: 100.00% Conservatives: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 49.86% Indels: 0
 Gaps: 6
 DB:
 US-10-037-417-46 (1-357) x AX342934 (1-537)
 QY 76 LeuIleAlaProSerTrpValLeuSerAlaAlaHisCysPheMetThrAsnGlyThrLeu 95
 Db 536 CTCATCGCCCTCTGGGTCTCTCGGTGCTCTCTGTCATGACGAATGGAGCGCTG 477
 QY 96 GluProAlaAlaGluTrpSerValLeuLeuGlyValHisSerGlnAspGlyProLeuAsp 115
 Db 476 GAGCCCGCGCGAGTGGTTCGTACTGCTGGCGGTGCATCTCCAGGAGCGGCGCCCTGGAC 417
 QY 116 GlyAlaHisThrArgAlaValAlaAlaIleValProAlaAsnTrpSerGlnValGlu 135
 Db 416 GGCGCGCACCGCGCAGTGGCGCGCATCGTGTGGCGGCACTACAGCAAGTGGAG 357
 QY 136 LeuGlyAlaAspLeuAlaLeuLeuArgLeuAlaSerProAlaSerLeuGlyProAlaVal 155
 Db 356 CTGGCGCGCGGACCTGGGCGCTGCTGGCTCACCGCCAGCTGGGCGCGCGCGTGT 297
 QY 156 TrpProValCysLeuProArgAlaSerHisArgPheValHisGlyThrAlaCysTrpAla 175
 Db 296 TGGCTGTCTGCTGCTGGCGCGCTGCACACGCTTCGTGACGCGGACCGCTTGGGCGC 237
 QY 176 ThrGlyTrpGlyAspValGlnGluAlaAspProLeuProLeuProTrpValLeuGlnGlu 195
 Db 236 ACCGCTGGGAGAGCTCCAGGAGCATCTCTGCGCTCTCCCTGGGTGCTACAGGAA 177
 QY 196 ValGluLeuArgLeuLeuGlyGluAlaThrCysGlnCysLeuTrpSerGlnProGlyPro 215
 Db 176 GTGGAGCTTAAGCTGCTGGCGAGCGCCCTGTCAATGCTCTACAGCCAGCGCGTCCC 117
 QY 216 PheAsnLeuThrLeuGlnIleLeuProGlyMetLeuCysAlaGlyTrpProGluGlyArg 235
 Db 116 TTCACCTCACTCTCAGATATTCAGGAGTGGTGTGTGGCTACCCAGGAGCGCGC 57
 QY 236 ArgAspThrCysGlnGlyAspSerGlyProLeuValCysGlnGluGlyGly 253
 Db 56 AGGAGACCTGCGAGGCTGACTCTGGGGGGCGCCCTGCTGTGTGAGGAGGCGCGC 3

RESULT 8
 BD125219
 LOCUS
 DEFINITION
 ACCESSION
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM
 Homo sapiens
 Homo sapiens
 Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 670)
 Ota, T., Nishikawa, T., Isogai, T., Hayashi, K., Ishii, S., Kawai, Y.,
 Wakamatsu, A., Sugiyama, T., Nagai, K., Kojima, S., Otsuki, T. and
 Koga, H.
 Primer for synthesizing full-length cDNA and use thereof
 Patent: JP 2002017375-A 650 22-JAN-2002;
 HELIX RESEARCH INSTITUTE
 OS Homo sapiens (human)
 PN JP 2002017375-A/650
 PD 22-JAN-2002
 PF 07-JUL-2000 JP 2000253172
 PI TOSHIO OTA, TETSUO NISHIKAWA, TAKAO ISOGAI, KOJI HAYASHI, SHIZUKO
 PI ISHII,
 PI YURI KAWAI, AI WAKAMATSU, TOMOYASU SUGIYAMA, KEIICHI NAGAI, PI


```

REFERENCE
AUTHORS      Xiao,Y. and Morozov,V.
TITLE        Regulation of human prostatic-like serine protease
JOURNAL      Patent: WO 0198467-A 3 27-DEC-2001;
              Bayer Aktiengesellschaft (DE)
FEATURES
source       Location/Qualifiers
            1..456
              /organism="Homo sapiens"
              /mol_type="unassigned DNA"
              /db_xref="taxon:9606"

ORIGIN
Alignment Scores:      6.98e-120      Length:      456
Pred. No.:            128.00          Matches:     128
Score:                100.00%          Conservative: 0
Percent Similarity:   100.00%          Mismatches:  0
Best Local Similarity: 100.00%          Indels:      0
Query Match:         35.85%            Gaps:       6
DB:

US-10-037-417-46 (1-357) x AX342936 (1-456)
QY 121 AlaValAlaAlaIleValValProAlaAsnTyrSerGlnValGluLeuGlyAlaAspLeu 140
DB 384 GCAGTGGCGCCATCGTGTGGCGGCCAACTACAGCAAGTGGAGCTGGGGCGGACCTG 325
QY 141 AlaLeuLeuArgLeuAlaSerProAlaSerLeuGlyProAlaValTyrProValCysLeu 160
DB 324 GCCTGTCTGGCGCTGCTCCACCCGCCAGCTGGGGCCCGCGTGGCTGTCTGCGCTG 265
QY 161 ProArgAlaSerHisArgPheValHisGlyThrAlaCysTyrAlaThrGlyTyrGlyAsp 180
DB 264 CCCCAGCGCTCACACCGCTTCGTGACGCGACCGCTGTGGGCCACCGCTGGGGAGAC 205
QY 181 ValGlnGluAlaAspProLeuProLeuProTyrValLeuGlnGluValGluLeuArgLeu 200
DB 204 GTCCAGGAGCGAGATCTCTGCTCTCTCCCTGCTTACAGGAGTGGAGCTAAGGCTG 145
QY 201 LeuGlyGluAlaThrCysGlnCysLeuTyrSerGlnProGlyProPheAsnLeuThrLeu 220
DB 144 CTGGCGGAGCCACCTGTCAATGTCTACAGCACCGCCGGTCCCTTCACCTCACTCTC 85
QY 221 GlnIleLeuProGlyMetLeuCysAlaGlyTyrProGluGlyValArgAspThrCysGln 240
DB 84 CAGATATTGCCAGGATGCTGTGTGTGCTACCCAGAGGCGCGCAGGACACCTGCCAG 25
QY 241 GlyAspSerGlyGlyProLeuVal 248
DB 24 GGTGACTCTGGGGGGCCCTGGTC 1

RESULT 11
AC009088/c 127769 bp DNA linear PRI 29-MAR-2003
LOCUS      Homo sapiens chromosome 16 clone RP11-388M20, complete sequence.
DEFINITION AC009088
ACCESSION  AC009088.9 GI:29366934
VERSION    HTG.
KEYWORDS   Homo sapiens (human)
SOURCE     Homo sapiens
           Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
           Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE  1 (bases 1 to 127769)
AUTHORS   DOE Joint Genome Institute, Stanford Human Genome Center and Los
           Alamos National Laboratory.
TITLE     Direct Submission
JOURNAL   Unpublished
REFERENCE  2 (bases 1 to 127769)
AUTHORS   DOE Joint Genome Institute.
TITLE     Direct Submission
JOURNAL   Submitted (03-AUG-1999) Production Sequencing Facility, DOE Joint
           Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
REFERENCE  3 (bases 1 to 127769)
AUTHORS   DOE Joint Genome Institute.

```

```

TITLE      Direct Submission
JOURNAL    Submitted (03-APR-2002) Production Sequencing Facility, DOE Joint
           Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
REFERENCE  4 (bases 1 to 127769)
AUTHORS   DOE Joint Genome Institute, Stanford Human Genome Center and Los
           Alamos National Laboratory.
TITLE      Direct Submission
JOURNAL    Submitted (18-MAR-2003) DOE Joint Genome Institute, 2800 Mitchell
           Drive, Walnut Creek, CA 94598, USA
REFERENCE  5 (bases 1 to 127769)
AUTHORS   DOE Joint Genome Institute, Stanford Human Genome Center and Los
           Alamos National Laboratory.
TITLE      Direct Submission
JOURNAL    Submitted (29-MAR-2003) DOE Joint Genome Institute, 2800 Mitchell
           Drive, Walnut Creek, CA 94598, USA
COMMENT    On Mar 29, 2003 this sequence version replaced gi:29029216.
           Draft Sequence produced by DOE Joint Genome Institute
           www.jgi.doe.gov
           Finishing Completed at Stanford Human Genome Center and Los Alamos
           National Laboratory
           www.sbgc.stanford.edu
           Quality: Phrap Quality >=40 99.7% of Sequence;
           Estimated Total Number of Errors is 0.2.
           NOTE: This insert is not the entire sequence of the clone (entire
           sequence is 233.4kb). It is clipped at the overlap with AC135050.
           The number of bases overlapped is 2575.
FEATURES
Location/Qualifiers
            1..127769
              /organism="Homo sapiens"
              /mol_type="genomic DNA"
              /db_xref="taxon:9606"
              /chromosome="16"
              /clone="RP11-388M20"
ORIGIN
Alignment Scores:      1.31e-82      Length:      127769
Pred. No.:            93.00          Matches:     93
Score:                100.00%          Conservative: 0
Percent Similarity:   100.00%          Mismatches:  0
Best Local Similarity: 100.00%          Indels:      0
Query Match:         26.05%            Gaps:       0
DB:

US-10-037-417-46 (1-357) x AC009088 (1-127769)
QY 92 AsnGlyThrLeuGluProAlaAlaGluTyrSerValLeuLeuGlyValHisSerGlnAsp 111
DB 31589 AATGGAGCGCTGGAGCCCGCGCGAGTGGTGGTACTGCTGGCGGTGCACCTCCAGGAC 31530
QY 112 GlyProLeuAspGlyAlaHisThrArgAlaValAlaAlaIleValValProAlaAsnTyr 131
DB 31529 GGGCCCCCTGGAGCGCGCGCACACCGCGCAGTGGCGCCCATCTGTGTGTGGGCACTAC 31470
QY 132 SerGlnValGluLeuGlyAlaAspLeuAlaLeuLeuArgLeuAlaSerProAlaSerLeu 151
DB 31469 AGCCAAGTGGAGCTGGCGCGCGACCTGGCCCTGCTGGCTTACCGCGCAGCCTG 31410
QY 152 GlyProAlaValTyrProValCysLeuProArgAlaSerHisArgPheValHisGlyThr 171
DB 31409 GGCCCCCGCGTGGCGCTGTCTGCTTCCCTGGCGCGCTCACCGCTTGTGTGACGGGACC 31350
QY 172 AlaCysTyrAlaThrGlyTyrGlyAspValGlnGluAla 184
DB 31349 GCCTGTGGGCCACCGCGCTGGGGAGACGTCCAGGAGGCA 31311

RESULT 12
AC135044 195476 bp DNA linear HTG 05-OCT-2002
LOCUS      Homo sapiens chromosome 16 clone CTD-2551B20, WORKING DRAFT
DEFINITION AC135044
ACCESSION  AC135044.1 GI:23505535
VERSION    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ACTIVEFIN.
KEYWORDS

```

SOURCE ORGANISM

Homo sapiens (human)
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.

REFERENCE

1 (bases 1 to 195476)
DOE Joint Genome Institute.

AUTHORS

Unpublished
Sequencing of Human Chromosome 16
2 (bases 1 to 195476)
DOE Joint Genome Institute.

TITLE

Submitted (05-OCT-2002) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA

JOURNAL

-----Genome Center
Center: Joint Genome Institute
Center Code: JGI
Web site: <http://www.jgi.doe.gov>

COMMENT

Project Information
Center Project Name: 809609
Center clone name: C17B-B1_2551B20

Summary Statistics

Consensus quality: 171229 bases at least Q40
Consensus quality: 182638 bases at least Q30
Consensus quality: 188095 bases at least Q20
Estimated insert size: 170000; agarose-fp estimation
Quality coverage: 7.63 in Q20 bases; agarose-fp estimation
Quality coverage: 6.7 in Q20 bases; sum-of-contigs estimation.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 22 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1 1202: contig of 1202 bp in length
1203 1302: gap of unknown length
1303 2468: contig of 1166 bp in length
2469 2568: gap of unknown length
2569 4077: contig of 1509 bp in length
4078 4177: gap of unknown length
4178 5772: contig of 1595 bp in length
5773 5872: gap of unknown length
5873 7722: contig of 1850 bp in length
7723 7822: gap of unknown length
7823 8958: contig of 1136 bp in length
8959 9058: gap of unknown length
9059 11051: contig of 1993 bp in length
11052 11151: gap of unknown length
11152 13641: contig of 2490 bp in length
13642 13741: gap of unknown length
13742 17738: contig of 3997 bp in length
17739 17838: gap of unknown length
17839 22030: contig of 4192 bp in length
22031 22130: gap of unknown length
22131 27002: contig of 4872 bp in length
27003 32819: contig of 5717 bp in length
32820 32919: gap of unknown length
32920 39003: contig of 6084 bp in length
39004 39103: gap of unknown length
39104 48370: contig of 9267 bp in length
48371 48470: gap of unknown length
48471 59721: contig of 11251 bp in length
59722 59821: gap of unknown length
59822 72404: contig of 12483 bp in length
72405 72406: gap of unknown length
72407 85349: contig of 12945 bp in length
85350 85449: gap of unknown length
85450 101872: contig of 16423 bp in length

* 101873 101972: gap of unknown length
* 101973 117076: contig of 15104 bp in length
* 117077 117176: gap of unknown length
* 117177 138973: contig of 21797 bp in length
* 138974 139073: gap of unknown length
* 139074 150336: contig of 11263 bp in length
* 150337 150436: gap of unknown length
* 150437 195476: contig of 45040 bp in length.

FEATURES

Location/Qualifiers
1..195476
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="16"
/clone="CTD-2551B20"
/clone_lib="Caltech human BAC library D"

ORIGIN

Alignment Scores:
Pred. No.: 1.79e-82 Length: 195476
Score: 93.00 Matches: 93
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 26.05% Indels: 0
DB: 2 Gaps: 0

US-10-037-417-46 (1-357) x AC135044 (1-195476)

Qy 92 AsnGlyThrLeuGluProAlaAlaGluTrpSerValLeuLeuGlyValHisSerGlnAsp 111
Db 27407 AATGGACGCTGGAGCCGCGCGGCGGAGTGGTGGTACTGTGGCGTGCACTCCAGGAC 27466
Qy 112 GlyProLeuAspGlyAlaHisThrArgAlaValAlaAlaValValProAlaAsnTyr 131
Db 27467 GGGCCCTCGACGCGCGCACACCCGCGCAGTGGCGCCATCGTGGTGGCGGCACTAC 27526
Qy 132 SerGlnValGluLeuGlyAlaAspLeuAlaLeuLeuArgLeuAlaSerProAlaSerLeu 151
Db 27527 AGCCACGTGGAGCTGGCGCGCGGCGGCGCTGGCGCTGGCGCTCACCCCGCGGCTG 27586
Qy 152 GlyProAlaValTrpProValCysLeuProArgAlaSerHisArgPheValHisGlyThr 171
Db 27587 GGGCCCGCGCTGGTGGCTGTGCTGCTGCGCGCGGCGCTCACACGCTTCTGTGACGGCACC 27646
Qy 172 AlaCysTrpAlaThrGlyTrpGlyAspValGlnGluAla 184
Db 27647 GCCTGCTGGCGCACCGGCTGGCGGAGACGTCAGAGGCA 27685

RESULT 13

AC106629/c 232119 bp DNA linear HTG 13-MAY-2003
LOCUS Rattus norvegicus clone CH230-21802, WORKING DRAFT SEQUENCE, 2
DEFINITION unordered pieces.

ACCESSION

AC106629.4 GI:30579402
HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
KEYWORDS Rattus norvegicus (Norway rat)
SOURCE Rattus norvegicus

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Euthera; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.

REFERENCE

1 (Bases 1 to 232119)
Muzny, D. Marie., Metzker, M. Lee., Abramson, S., Adams, C., Alder, J.,
Allen, C., Allen, H., Alsbrooks, S., Amin, A., Anguiano, D.,
Anyalebechi, V., Aoyagi, A., Ayodeji, M., Baca, E., Baden, H.,
Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benahmed, F.,
Biswal, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M.,
Bryant, N., Buhay, C., Burch, P., Burrell, K., Calderon, E.,
Cardenas, V., Carter, K., Cavazos, I., Ceasar, H., Center, A.,
Chacko, J., Chavez, D., Chen, G., Chen, Y., Chen, Z., Chu, J.,
Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L.,
Davila, M. L., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D.,
Delgado, O., Denson, S., Deramo, C., Ding, Y., Dinh, H., Divya, K.,

Anayalebechi, V., Aoyagi, A., Ayodeji, M., Baca, E., Baden, H.,
 Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benahmed, F.,
 Biswal, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M.,
 Bryant, N., Buhay, C., Burch, P., Burrell, K., Calderon, E.,
 Cardenas, V., Carter, K., Cavazos, I., Caesar, H., Center, A.,
 Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Y., Chen, Z., Chu, J.,
 Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L.,
 Davila, M.L., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D.,
 Delgado, O., Denson, S., Deramo, C., Ding, Y., Dinh, H., Divya, K.,
 Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Eaves, K.,
 Egan, A., Escotto, M., Eugene, C., Evans, C.A., Falls, T., Fan, G.,
 Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, P.,
 Fraser, C.M., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Garza, M.,
 Gregeorgis, E., Geer, K., Gail, R., Grady, M., Guerra, T., Guevara, W.,
 Gunaratne, P., Haaland, W., Hamil, C., Hamilton, C., Hamilton, K.,
 Harvey, Y., Havlak, P., Hawes, A., Henderson, N., Hernandez, J.,
 Hernandez, R., Hines, S., Hladun, S.L., Hodgson, A., Hoques, M.,
 Hollins, B., Howells, S., Hulyk, S., Hume, J., Idlebird, D., Jackson, A.,
 Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A.,
 Karpach, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C.,
 Kowis, C., Kraft, C.L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J.,
 Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J.,
 Lorensuhewa, L., Louisseg, H., Lozado, R.J., Lu, X., Ma, J.,
 Maheshwari, M., Mahindaratne, M., Mahmoud, M., Malloy, K., Mangum, A.,
 Mangum, B., Mapua, P., Martin, K., Martin, R., Martinez, E.,
 Mahoney, S., McLeod, M.P., McNeill, T.Z., Meenen, E.,
 Milosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S.,
 Morgan, M., Morris, K., Morris, S., Munidasa, M., Murphy, M., Nair, L.,
 Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S., Nsamenang, A.,
 Nwakenleh, O., Okwuonu, G., Olarunpunsagcon, A., Pal, S., Parks, K.,
 Pasternak, S., Paul, H., Perez, A., Perez, L., Pfannkuch, C.,
 Plopper, F., Poindexter, A., Popovic, D., Primus, E., Pu, L.-L.,
 Puzo, M., Quiroz, J., Rachlin, E., Reeves, K., Regier, M.A., Reigh, R.,
 Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F.,
 Rives, C., Rodkey, I., Rojars, A., Rose, M., Rose, R., Ruiz, S.J., Shen, H.,
 Sanders, W., Savery, G., Scherer, S., Scott, G., Shatsman, S., Shen, H.,
 Shetty, J., Shvartsbeyn, A., Sisson, I., Sitter, C.D., Smajs, D.,
 Sneed, A., Sodergren, E., Song, X.-Z., Sorelle, R., Sosa, J.,
 Steimle, M., Strong, R., Sutton, A., Svatek, A., Taber, P., Taylor, C.,
 Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Usmani, K.,
 Valas, R., Vera, V., Villanueva, D., Waldron, L., Walker, B., Wang, J.,
 Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F.,
 Williams, G., Willson, R., Wleczyk, R., Wooden, H., Worley, K.,
 Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, D., Yoon, V.,
 Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von
 Niederhausern, A., Weiss, R., Smith, D.R., Holt, R.A., Smith, H.O.,
 Weinstock, G. and Gibbs, R.A.
 Direct Submission
 2 (bases 1 to 244161)
 Worely, K.C.
 Direct Submission
 Submitted (08-APR-2002) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA
 3 (bases 1 to 244161)
 Rat Genome Sequencing Consortium.
 Direct Submission
 Submitted (15-NOV-2002) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA
 On Nov 15, 2002 this sequence version replaced gi:23267402.
 The sequence in this assembly is a combination of BAC based reads
 and whole genome shotgun sequencing reads assembled using Atlas
 (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
 in the feature table below represents a scaffold in the Atlas
 assembly (a 'contig-scaffold'). Within each contig-scaffold,
 individual sequence contigs are ordered and oriented, and separated
 by sized gaps filled with Ns to the estimated size. The sequence
 may extend beyond the ends of the clone and there may be sequence
 contigs within a contig-scaffold that consist entirely of whole
 genome shotgun sequence reads. Both end sequences and whole genome
 shotgun sequence only contigs will be indicated in the feature

table.

----- Genome Center
 Center: Baylor College of Medicine
 Center code: BCM
 Web site: <http://www.hgsc.bcm.tmc.edu/>
 Contact: hgsc-help@bcm.tmc.edu
 ----- Project Information
 Center project name: GWR
 Center clone name: CH230-20P9
 ----- Summary Statistics
 Assembly program: Phrap; version 0.990329
 Consensus quality: 223430 bases at least Q40
 Consensus quality: 226412 bases at least Q30
 Consensus quality: 228516 bases at least Q20
 Estimated insert size: 226458; sum-of-contigs estimation
 Quality coverage: 6x in Q20 bases; sum-of-contigs estimation

 * NOTE: Estimated insert size may differ from sequence length
 * (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 5 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.
 * 1 25619: contig of 25619 bp in length
 * 25620 25719: gap of unknown length
 * 137872: contig of 112153 bp in length
 * 137873 137972: gap of unknown length
 * 137973 240780: contig of 102818 bp in length
 * 240791 240890: gap of unknown length
 * 240891 241970: contig of 1080 bp in length
 * 241971 242070: gap of unknown length
 * 242071 244161: contig of 2091 bp in length.
 ----- Location/Qualifiers
 1. 244161
 /organism="Rattus norvegicus"
 /mol_type="genomic DNA"
 /db_xref="taxon:10116"
 /clone="CH230-20P9"
 19835..20652
 /note="clone_boundary
 clone_end:Sp6
 site:
 end_sequence:BH342986"
 21143..23721
 /note="wgs contig"
 23772..25619
 /note="wgs contig"
 103292..103777
 /note="clone_boundary
 clone_end:T7
 site:
 end_sequence:BH342985"
 129838..131159
 /note="wgs end_extension
 clone_end:T7"
 136097..137872
 /note="wgs end_extension
 clone_end:T7"

 ORIGIN
 Alignment Scores:
 Pred. No.: 1,348-23 Length: 244161
 Score: 35.00 Matches: 35
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 9.80% Indels: 0
 DB: 2 Gaps: 0
 US-10-037-417-46 (1-357) x ACL11710 (1-244161)

Qy

55

ProGlyThrTrpProTrpGlnValSerLeuHisHisGlyGlyGlyHisIleCysGlyGly 74

Db

129868

CCTGGCACTTGGCGTGGCAGTGAGCTGCATCACGTTGGAGCCACATCTGCGGGGC 129927

Qy

75

SerLeuIleAlaProSerTrpValLeuSerAlaAlaHisCysPhe 89

Db

129928

TCCCTCATCGCCCTTCTCTGGGTCTCTCCGCTGCTCAGTCTTTC 129972

AC093520

168064 bp

DNA

linear

PRI 19-MAR-2003

LOCUS

AC093520/c

Homo sapiens chromosome 16 clone RP11-120K18, complete sequence.

DEFINITION

AC093520

ACCESSION

AC093520

VERSION

AC093520.4

GI:29124069

HTG.

KEYWORDS

HTG.

SOURCE

Homo sapiens (human)

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE

1 (bases 1 to 168064)

AUTHORS

DOE Joint Genome Institute, Stanford Human Genome Center and Los Alamos National Laboratory.

TITLE

Direct Submission

JOURNAL

Unpublished

REFERENCE

2 (bases 1 to 168064)

AUTHORS

DOE Joint Genome Institute.

TITLE

Direct Submission

JOURNAL

Submitted (31-AUG-2001) Production Sequencing Facility, DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA

REFERENCE

3 (bases 1 to 168064)

AUTHORS

DOE Joint Genome Institute.

TITLE

Direct Submission

JOURNAL

Submitted (13-JUN-2002) Production Sequencing Facility, DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA

REFERENCE

4 (bases 1 to 168064)

AUTHORS

DOE Joint Genome Institute.

TITLE

Direct Submission

JOURNAL

Submitted (07-JUL-2002) Production Sequencing Facility, DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA

REFERENCE

5 (bases 1 to 168064)

AUTHORS

DOE Joint Genome Institute, Stanford Human Genome Center and Los Alamos National Laboratory.

TITLE

Direct Submission

JOURNAL

Submitted (19-MAR-2003) DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA

COMMENT

On Mar 19, 2003 this sequence version replaced gi:21702833.
Draft Sequence Produced by DOE Joint Genome Institute
www.igi.doe.gov
Finishing Completed at Stanford Human Genome Center and Los Alamos National Laboratory
www.sngc.stanford.edu
Quality: Phrap Quality >=40 100% of Sequence;
Estimated Total Number of Errors is 0.

Qy

312

HisGlnThrAlaPheLeuAepSerAlaArgIleLeuLeuArgProLeuSerHisIleSer 331

Db

141592

CATCAACGGCAATTCCTGGATTCTGCCAGAAATCCTTTTGAGGCCCTTGCCCATATCA 141533

Qy

332

ValGlyValSerThrGlyThrLysSer 340

Db

141532

GTAGGAGTCTCAACTGGACCAAAAGC 141506

Search completed: February 25, 2004, 23:28:10

Job time : 3995 secs

FEATURES

source

1. 168064

/organism="Homo sapiens"

/mol_type="genomic DNA"

/db_xref="taxon:9606"

/chromosome="16"

/clones="RP11-120K18"

ORIGIN

Alignment Scores:

Pred. No.: 1.23e-17

Score: 29.00

Percent Similarity: 100.00%

Best Local Similarity: 100.00%

Query Match: 8.12%

DB: 9

Matches: 29

Conservative: 0

Mismatches: 0

Indels: 0

Gaps: 0

US-10-037-417-46 (1-357) x AC093520 (1-168064)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: February 25, 2004, 18:01:15 ; Search time 419 Seconds
(without alignments)

3619.586 Million cell updates/sec

Title: US-10-037-417-46

Perfect score: 357

Sequence: 1 MAQGVLFGLGAVANSDS.....TKSLVLPWLSPLHLLGLWG 357

Scoring table:

OLIGO
Xgapop 60.0 , Xgapext 60.0
Ygapop 60.0 , Ygapext 60.0
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 3373863 seqs, 2124099041 residues

Word size: 1

Total number of hits satisfying chosen parameters: 6745146

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlp
-Q=/cgn2_1/USPTO_spool_p/US10037417/runat_24022004_101108_18072/app_query.fasta_1.519
-DB=N Geneseq 29Jan04 -QMT=fastap -SUFFIX=olip2n.rng -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=oligo -TRANS=human40.cdi
-LIST=45 -DOALIGN=200 -THR SCORE=quality -THR MIN=1 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10037417 @CGN 1_1_885 @runat_24022004_101108_18072 -NCPU=6 -ICPU=3
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database : N Geneseq 29Jan04:*

- 1: geneseqn1980s:*
- 2: geneseqn1990s:*
- 3: geneseqn2000s:*
- 4: geneseqn2001as:*
- 5: geneseqn2001bs:*
- 6: geneseqn2002as:*
- 7: geneseqn2003as:*
- 8: geneseqn2003bs:*
- 9: geneseqn2003cs:*
- 10: geneseqn2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	357	100.0	1102	6	Abq93901 Human pro
2	357	100.0	1102	6	Abq93902 Human pro
3	272	76.2	2457	6	Abk31795 DNA encod
4	272	76.2	2681	6	Abq75956 Human PMW
5	261	73.1	2810	4	Aak94500 Human ful
6	178	49.9	537	6	Abk13565 Human pro
7	132	37.0	597	5	Aas69040 DNA encod
8	132	37.0	670	4	Aak92190 Human cDN

c	9	132	37.0	670	4	AAK93456	Human cDN
	10	128	35.9	456	6	ABK13566	Human pro
	11	52	14.6	2298	5	AAH71079	DNA encod
	12	45	12.6	768	4	AAH78206	Nucleotid
c	13	17	4.8	8999	5	ABA21236	Human ner
	14	16	4.5	1161	6	ABN85393	Human NOV
	15	16	4.5	1726	6	ABN85392	Human NOV
	16	16	4.5	1733	7	ABT31936	Human bre
	17	16	4.5	1796	5	AAF98698	Human ova
	18	16	4.5	1834	6	ABS76501	cDNA enco
	19	16	4.5	1834	6	ABL67949	Ovary can
	20	16	4.5	1834	6	ABL68512	Kidney ca
	21	16	4.5	1834	6	ABL68248	Kidney ca
	22	16	4.5	1834	6	ABK12241	cDNA enco
	23	16	4.5	1834	6	ABN95716	Gene #221
	24	16	4.5	1835	5	AAF98720	Human lat
	25	16	4.5	3382	6	ABZ35336	Human gen
	26	13	3.6	307	7	ADA05695	Human NOV
	27	13	3.6	358	7	ADA05693	Human NOV
	28	13	3.6	436	7	ADA05699	Human NOV
	29	13	3.6	445	7	ADA05691	Human NOV
	30	13	3.6	484	7	ADA05687	Human NOV
c	31	13	3.6	506	6	ABK30302	Human G-p
	32	13	3.6	711	7	ADA05020	Human pro
	33	13	3.6	783	9	ADD09102	Human kal
	34	13	3.6	792	2	AAH80905	Human cDN
	35	13	3.6	809	5	AAH87560	DNA encod
	36	13	3.6	831	5	AAH85946	DNA encod
	37	13	3.6	845	7	ADA05701	Human NOV
	38	13	3.6	868	3	AAZ44182	Human neu
	39	13	3.6	905	2	AAZ11029	Human ser
	40	13	3.6	942	2	AAZ11030	Human ser
	41	13	3.6	944	2	AAV84052	Nucleic a
	42	13	3.6	988	6	ABZ35510	Human gen
	43	13	3.6	994	2	AAH87155	Human pro
	44	13	3.6	994	6	ABZ35501	Human gen
	45	13	3.6	999	2	AAZ34225	Human PRO

ALIGNMENTS

RESULT 1
ABQ93901
ID ABQ93901 standard; DNA; 1102 BP.
XX AC ABQ93901;
XX AC ABQ93901;
DT 01-NOV-2002 (first entry)
DE Human prostatin precursor-like NOV14a DNA, SEQ ID NO:43.
KW Human; NOVX; neurological disorder; Alzheimer's disease;
KW Huntington's disease; Parkinson's disease; pain; behavioural disorder;
KW addition; tubercous sclerosis; cancer; immune disorder; allergy;
KW autoimmune disease; myasthenia gravis; asthma; arthritis; diabetes;
KW thyroiditis; cardiovascular disease; hypertension; reproductive disorder;
KW endometriosis; incontinence; psoriasis; scleroderma; alopecia; ulcer;
KW pancreatic cirrhosis; glomerular endotheliosis; bacterial infection;
KW polycystic kidney disease; endocrine disorder; obesity; cardiomyopathy;
KW atherosclerosis; cell signal processing-related disorder;
KW metabolic pathway regulation disorder; cytostatic; neuroprotective;
KW antiinflammatory; immunosuppressive; analgesic; antithrombotic;
KW dermatological; antibacterial; antiarthritic; hepatotropic; neurogenesis;
KW differentiation; proliferation; motility; haematopoiesis; wound healing;
KW angiogenesis; forensic biology; transgenic animal; drug screening;
KW gene therapy; NOV14a; prostatin precursor-like; chromosome 16; gene; ds.

Homo sapiens.

Location/Qualifiers
1..18
/*tag= a
19..1050
CDS

Db 919 ACCAGCCAGAGACCCAGTCAGATTCTTACATCAAGGCACTTCCTGGATTCTGCC 978
QY 321 ArgIleLeuLeuArgProLeuSerHisIleSerValGlyValSerThrGlyThrIleSer 340
Db 979 AGAATCCTTTTGAGGCCCTGTCTCCCATATATCATAGTAGAGTCTCAACTGGGACCAAAAGC 1038
QY 341 LeuValLeuProTrpLeuSerProHisSerLeuLeuGlyLeuTrpGlyPhe 357
Db 1039 CTTGTCTCCCTCCCTGGCTCTCTCCACACTCTCTCTGGGCCCTCTGGGGGTTC 1089
RESULT 2
ABQ93902
ID ABQ93902 standard; DNA; 1102 BP.
XX AC ABQ93902;
XX 01-NOV-2002 (first entry)
XX Human prostatin precursor-like NOV14b DNA, SEQ ID NO:45.
XX Human; NOVX; neurological disorder; Alzheimer's disease;
KW Huntington's disease; Parkinson's disease; pain; behavioural disorder;
KW addition; tuberous sclerosis; cancer; immune disorder; allergy;
KW autoimmune disease; myasthenia gravis; asthma; arthritis; diabetes;
KW thyroiditis; cardiovascular disease; hypertension; reproductive disorder;
KW endometriosis; incontinence; psoriasis; scleroderma; alopecia; ulcer;
KW pancreatic disease; glomerular endotheliosis; bacterial infection;
KW polycystic kidney disease; endocrine disorder; obesity; cardiomyopathy;
KW atherosclerosis; cell signal processing-related disorder;
KW metabolic pathway regulation disorder; cytostatic; neuroprotective;
KW antiinflammatory; immunosuppressive; analgesic; antihypertensive;
KW dermatological; antibacterial; antiarthritic; hepatotropic; neurogenesis;
KW differentiation; proliferation; motility; haematopoiesis; wound healing;
KW angiogenesis; forensic biology; transgenic animal; drug screening;
KW gene therapy; NOV14b; prostatin precursor-like; gene; ds.
XX Homo sapiens.
XX Key Location/Qualifiers
FH 5'UTR 1..18
FT /*tag= a
FT 19..1092
FT /*tag= b
FT /*product= "NOV14b"
FT 1093..1102
FT /*tag= c
XX WO200253742-A2.
XX 11-JUL-2002.
XX 07-JAN-2002; 2002WO-US000375.
XX 05-JAN-2001; 2001US-0260018P.
XX 08-JAN-2001; 2001US-0260360P.
XX 28-FEB-2001; 2001US-0272411P.
XX 02-MAR-2001; 2001US-0272817P.
XX 05-JUL-2001; 2001US-0303231P.
XX 12-JUL-2001; 2001US-0305060P.
XX 10-SEP-2001; 2001US-0318405P.
XX 12-SEP-2001; 2001US-0318700P.
XX 04-JAN-2002; 2002US-00037417.
XX (CURA-) CURAGEN CORP.
XX Kekuda R, Alsbrook UP, Tchernev VT, Liu X, Spytek KA;
PI Pattarajan M, Grosse WM, Lepley DM, Burgess CE, Vernet CM, Li L;
PI Gorman L, Edinger S, Sciore P, Eilerman X, Malyankar U;
PI Rothenberg M, Stone D, Boldog F, Guo X, Shenoy S, Anderson D;
PI Padigaru M, Taupier RU, Miller CE, Eisen A;
XX WPI; 2002-583619/62.

DR P-PSDB; AB09524.
XX Novel polypeptides and nucleic acids homologous to transmembrane
PT receptor, thymosin, neuromodulin-like family of proteins for diagnosing,
PT treating cancer, atherosclerosis, neurological, skin and autoimmune
XX disorders.
PS Claim 9a; Page 143; 323pp; English.
XX The invention relates to 24 novel human proteins designated NOV1-NOV14
CC (AB09501-AB09524), collectively referred to as NOVX proteins, and
CC nucleic acids encoding them (AB093879-AB093902). NOVX proteins and
CC nucleotides are useful in the treatment, diagnosis or prevention of NOVX-
CC associated disorders or in the manufacture of a medicament for treating
CC such disorders, with specific applications described for each of the 24
CC NOVX proteins, based on their homology to known proteins. Various
CC disorders are associated with NOVX proteins including neurological
CC disorders (e.g., Alzheimer's, Huntington's and Parkinson's diseases),
CC pain, behavioural disorders, addiction, tuberculous sclerosis, cancers
CC (e.g., colorectal cancer, leukaemia and osteosarcoma), immune disorders
CC (e.g., allergies and autoimmune diseases), myasthenia gravis, asthma,
CC various forms of arthritis, diabetes, thyroiditis, cardiovascular disease
CC (e.g., hypertension), reproductive disorders, endometriosis,
CC incontinence, psoriasis, scleroderma, alopecia, ulcers, pancreatitis,
CC cirrhosis, glomerular endotheliosis, polycystic kidney disease, endocrine
CC disorders, obesity, bacterial infections and particularly cardiomyopathy,
CC atherosclerosis, cell signal processing-related disorders and disorders
CC of metabolic pathway regulation. NOVX nucleic acids and polypeptides may
CC be used to identify cellular receptors or downstream effectors which
CC binds to a NOVX protein, and are also useful as targets for the
CC identification of small molecules that modulate or inhibit processes such
CC as neurogenesis, cell differentiation, cell motility, cellular
CC proliferation, haematopoiesis, wound healing and angiogenesis. NOVX
CC nucleic acid sequences can be used to identify a cell or tissue type and
CC are useful as a source of primers or probes for forensic biology and for
CC identifying and cloning NOVX homologues in other cell types. Cells
CC comprising NOVX nucleic acids are useful for producing non-human
CC transgenic animals which are useful for studying the function and
CC activity of NOVX proteins and for identifying and evaluating modulators
CC of NOVX activity. The present sequence represents DNA encoding the
CC prostatin precursor-like protein NOV14b
XX SQ Sequence 1102 BP; 162 A; 371 C; 353 G; 216 T; 0 U; 0 Other;
Alignment Scores:
Pred. No.: 0 Length: 1102
Score: 357.00 Matches: 357
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: Gaps: 6
US-10-037-417-46 (1-357) x ABQ93902 (1-1102)
QY 1 MetaLaGlnLysGlyValLeuGlyProGlyGlnLeuGlyAlaValAlaAsnSerAspSer 20
Db 19 ATGGCCAGAGAGGGGTCTGGGGCTTGGGAGCTGGGGCTGTGGCCAAATTCCTACTCA 78
QY 21 TyrSerLeuTyrGlyLeuValProSerGlyProAlaArgGlyProProTyrCysGlyArg 40
Db 79 TACTCACTTTACGGGTGTGGTCCGTCGGAGACCGCTAGGGGGCCCCCGTACTGGGGGCGC 138
QY 41 ProGluProSerAlaArgIleValGlyGlySerAsnAlaGlnProGlyThrTrpProTrp 60
Db 139 CCTGAGCCCTCGGCCGCAATCGTGGGGGGCTCAACCGCCAGCCGGGACCTTGGCTTGG 198
QY 61 GlnValSerLeuHisHisGlyGlyGlyHisIleCysGlyGlySerLeuLeuAlaProSer 80
Db 199 CAACTGAGCTGCTGCACCATCGAGGTGGCCACATCTGGGGGGCTCCCTCATCGCCCTCC 258
QY 81 TrpValLeuSerAlaAlaHisCysPheMetThrAsnGlyThrLeuGluProAlaAlaGlu 100
Db 259 TGGGTCTCTCTCGGCTGCTCACTGTGTTTCATGACGAATGGGACGTTGGAGCCCGCGCGCAG 318

Thu Feb 26 13:50:24 2004

us-10-037-417-46.olip2n.rng

QY 101 TrpSerValLeuLeuGlyValHisSerGlnAspGlyProLeuAspGlyAlaHisThrArg 120
 Db 319 TGGTCGGTACTGCTGGGGCTGCTACCTCCAGACGGGGCCCTGGAGCGCGGCACACCCGC 378
 QY 121 AlaValAlaIleValValProAlaAsnTyrSerGlnValGluLeuGlyAlaAspLeu 140
 Db 379 GCAGTGGCCGCCATCGTGGTGGCGGCCACTACAGCCAGGTGGAGCTGGGGCGCGAGCTG 438
 QY 141 AlaLeuLeuArgLeuAlaSerProAlaSerLeuGlyProAlaValTrrProValCysLeu 160
 Db 439 GCCCTGCTGGCGCTGCTCCAGCCAGCTGGGGCCCGCCGCTGGCCCTGCTGGCTG 498
 QY 161 ProArgAlaSerHisArgPheValHisGlyThrAlaCysTrrAlaThrGlyTrrGlyAsp 180
 Db 499 CCCCCTGGCTACACCGCTGCTGGTGGCGCCAGCGCTGCTGGCCACCGCTGGGGAGAC 558
 QY 181 ValGlnGluAlaAspProLeuProTrrValLeuGlnGluValGluLeuArgLeu 200
 Db 559 GTCCAGAGGCGAGTCTCTGCTCTCCCTGGGTGCTACAGGAAGTGGAGCTAAGGCTG 618
 QY 201 LeuGlyGluAlaThrCysGlnCysLeuTrrSerGlnProGlyProPheAsnLeuThrLeu 220
 Db 619 CTGGCGAGGCCACCTGCTCAATGCTCTACAGCCAGCCCGGCTCCCTTCAACCTCACTCTC 678
 QY 221 GlnIleLeuProGlyMetLeuCysAlaGlyTrrProGluGlyArgArgAspThrCysGln 240
 Db 679 CAGATATTGCCAGGGAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 738
 QY 241 GlyAspSerGlyGlyProLeuValCysGluGluGlyGlyArgTrrPheGlnAlaGlyIle 260
 Db 739 GGTGACTCTGGGGGGCCCTGCTGCTGAGGAAGCGCGGCTGGTTCAGGCGAGGAATC 798
 QY 261 ThrSerPheGlyPheGlyCysGlyArgArgAsnArgProGlyValPheThrAlaValAla 280
 Db 799 ACCAGCTTTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 858
 QY 281 ThrTyrGluAlaTrrPheArgGluGlnValMetGlySerGluProGlyProAlaPhePro 300
 Db 859 ACCTATGAGGCATGGATACGGGAGCAGGTGATGGGTTCAGAGCTGGGGCTGGCTTCCC 918
 QY 301 ThrGlnProGlnTrrGlnSerAspCysLeuHisGlnThrAlaPheLeuAspSerAla 320
 Db 919 ACCAGCCCCAGAGACCCAGTCAGATTGTTTACATCAACAGCGCATTCCTGGATTCTGCC 978
 QY 321 ArgIleLeuLeuArgProLeuSerHisIleSerValGlyValSerThrGlyThrLysSer 340
 Db 979 AGAATCCTTTTGGGGCCCTTGTCCCATATATCATAGTAGAGTCTCACTGGGACCAAGC 1038
 QY 341 LeuValLeuProTrrLeuSerProHisSerLeuLeuGlyLeuTrrGlyPhe 357
 Db 1039 CTGTGCTCTCCCTGGCTCTCTCCACACTCTCTCTGGGGCTCTGGGGGTTT 1089
 RESULT 3
 ID ABK31795
 XX ABK31795 standard; DNA; 2457 BP.
 XX AC ABK31795;
 XX DT 23-APR-2002 (first entry)
 XX DE DNA encoding novel human protease #52.
 XX KW Human; protease; cancer; immune-related disorder; cardiovascular disease;
 KW neuronal-associated disease; metabolic disorder; inflammatory disorder;
 KW nervous system disorder; sexual dysfunction; pain; mood disorder;
 KW hypertension; psychotic disorder; neurological disorder; dyskinesia;
 KW viral infection; human immunodeficiency virus; HIV; non-viral infection;
 KW ocular disease; cytostatic; gene; ds.
 XX OS Homo sapiens.
 XX PN W0200200860-A2.

XX 03-JAN-2002.
 XX 26-JUN-2001; 2001WO-US020171.
 XX 26-JUN-2000; 2000US-0214047P.
 XX (SUGE-) SUGEN INC.
 XX Plowman G, Whyte D, Sudarsanam S, Manning G, Caenepeel S;
 XX Charyczak G;
 DR WPI; 2002-139913/18.
 XX P-PSDB; AAU82753.
 XX Nucleic acids encoding novel human proteases, useful for useful for
 XX treating diseases and disorders such as cancers, immune-related diseases
 XX and disorders, cardiovascular disease (e.g. restenosis) and inflammatory
 XX disorders.
 XX Claim 26; Fig 18S-TT; 313pp; English.
 XX The present invention relates to the isolation of novel human proteases,
 XX and the nucleic acids encoding them. The sequences of the invention are
 XX useful for treating diseases and disorders such as cancers (e.g. breast,
 XX colon, lung), immune-related diseases and disorders (e.g. inflammatory
 XX diseases and asthma), cardiovascular diseases (e.g. restenosis and
 XX coronary thrombosis), brain or neuronal inflammatory disorders (e.g.
 XX disorders (e.g. diabetes, obesity), inflammatory disorders (e.g.
 XX rheumatoid arthritis and psoriasis), central or peripheral nervous system
 XX diseases, migraines, pain, sexual dysfunction, mood disorders, attention
 XX disorders, neurological disorders, hypotension, hypertension, psychotic
 XX disease) and dyskinesias. The nucleic acids and polypeptides are also
 XX useful for treating viral infections caused by human immunodeficiency
 XX virus (HIV), and non-viral infections such as ocular disease (e.g.
 XX glaucoma) and macular degeneration. ABK31744-ABK31802 represent DNA
 XX sequences encoding for the novel human proteases of the invention
 XX SQ Sequence 2457 BP; 377 A; 851 C; 797 G; 432 T; 0 U; 0 Other;
 Alignment Scores:
 Pred. No.: 7, 48e-240 Length: 2457
 Score: 272.00 Matches: 272
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 76.19% Indels: 0
 DB: 6 Gaps: 0
 US-10-037-417-46 (1-357) x ABK31795 (1-2457)
 QY 38 CysGlyArgProGluProSerAlaArgIleValGlyGlySerAsnAlaGlnProGlyThr 57
 Db 112 TGGCGGGCGCCCTGAGCCCTCGGCCCGCATCTGGGGGGGCTCAAAACCGCGACCGGGGACC 171
 QY 58 TrrProTrrGlnValSerLeuHisHisGlyGlyGlyHisIleCysGlyGlySerLeu 77
 Db 172 TGGCTTGGCAAGTGGAGCTTGCACCATGGAGGTGGCCACATCTCGGGGGCTCCCTCATC 231
 QY 78 AlaProSerTrrValLeuSerAlaAlaHisCysPheMetThrAsnGlyThrLeuGluPro 97
 Db 232 GCCCCTCTGGGTCTCTCCGCTGCTCACTGTTTCATGACGAATGGAGCGCTGGAGCCC 291
 QY 98 AlaAlaGluTrrSerValLeuLeuGlyValHisSerGlnAspGlyProLeuAspGlyAla 117
 Db 292 GCGGCCGAGTGGTGGTACTCTGGGGCGTGCATCCAGGACGGGGCCCTTGGAGCGGCG 351
 QY 118 HisThrArgAlaValAlaAlaIleValValProAlaAsnTrrSerGlnValGluLeuGly 137
 Db 352 CACACCCGCGAGTGGCGCCCATCTGTGTGCCGCCAACTACAGCAAGTGGAGCTGGGC 411
 QY 138 AlaAspLeuAlaLeuArgLeuAlaSerProAlaSerLeuGlyProAlaValTrrPro 157

Db 412 GCCGACCTGGCCCTGCTGCGCCTGACCCCGCCAGCCTGGGCCCCGCGTGTGGCCT 471
Qy 158 ValCysLeuProAlaSerHisArgPheValHisGlyThrAlaCysTrpAlaThrGly 177
Db 472 GTCTGCTGCCCCGCGCTACACCGCTTCGTGACGGCACCCTGCTGGGCCACCGGC 531
Qy 178 TrpGlyAspValGlnGluAlaAspProLeuProLeuProTrpValLeuGlnGluValGlu 197
Db 532 TGGGGAGACGTCACAGGAGGACAGATCCCTGCTGCTCCCTGCGTGTACAGGAAGTGGAG 591
Qy 198 LeuArgLeuLeuGlyGluAlaThrCysGlnCysLeuTyrSerGlnProGlyProPheAsn 217
Db 592 CTAAAGGCTGTGGCGGAGGCCACCTGTCAATGCTCTACAGCCAGCCCGTCCCTTCAAC 651
Qy 218 LeuThrLeuGlnLeuLeuProGlyMetLeuCysAlaGlyTyrProGluGlyArgArgAsp 237
Db 652 CTCACCTCCAGATATTGCCAGGGATGCTGTGTCTGCTACCCAGAGGGCCGAGGGAC 711
Qy 238 ThrCysGlnGlyAspSerGlyGlyProLeuValCysGluGluGlyGlyArgTrpPheGln 257
Db 712 ACCTGCCAGGGTGACTCTGGGGGGCCCCCTGCTGTGTGAGGAAGCGCGCTGTTCAG 771
Qy 258 AlaGlyIleThrSerPheGlyPheGlyCysGlyArgArgAsnArgProGlyValPheThr 277
Db 772 GCAGGAATCACAGCTTGGCTTTGGCTGTGTGGAGCGAAGAACCGCCCTGGAGTTTCACT 831
Qy 278 AlaValAlaThrTyrGluAlaTrpIleArgGluGlnValMetGlySerGluProGlyPro 297
Db 832 GCTGTGGCTACCTATGAGGCATGGATACGGGAGCAGGTGATGGGTTCAGAGCCTGGGCT 891
Qy 298 AlaPheProThrGlnProGlnLysThrGlnSerAsp 309
Db 892 GCCTTTCCACCCAGCCCGCCAGAACCCAGTCAGAT 927
RESULT 4
ABQ75956
ID ABQ75956 standard; cDNA; 2681 BP.
AC ABQ75956;
XX
DT 17-OCT-2002 (first entry)
DE Human PMMM encoding sequence Incyte ID 2751509CB1.
KW Human; PMMM; protein modification and maintenance molecule;
KW anticonvulsant; neuroprotective; nootropic; cytostatic; antipsoriatic;
KW antiasthmatic; dermatological; antidiabetic; antiparkinsonian;
KW antianemic; antiinflammatory; antiulcer; antianginal; cardiant;
KW hepatotropic; osteopathic; antiemetic; antipyretic; virucide;
KW antibacterial; fungicide; gastrointestinal; antidiarrheic; laxative;
KW haemostatic; anti-HIV; thrombolytic; anticoagulant; gynaecological;
KW cardiovascular; antiarteriosclerotic; hypotensive; vasotropic;
KW antitumour; antirheumatic; immunosuppressive; anti allergic; antithyroid;
KW nephrotropic; antigout; thyromimetic; antiarthritic; uropathic;
KW ophthalmological; antiparasitic; tranquiliser; vulnerary; keratolytic;
KW auditory; antiseborrheic; antidepressant; neuroleptic; antiinfertility;
KW antelmintic; protozoacide; Crohn's disease; hypertension; autoimmune;
KW inflammatory; anaemia; cell proliferative; developmental; epithelial;
KW scabies; neurological; Alzheimer's disease; reproductive;
KW ectopic pregnancy; gene therapy; vaccine; disorder; prostasin; gene; ss.
XX Homo sapiens.
OS
XX
FH Key Location/Qualifiers
CDS 1..2457
FT /*tag= a
FT /product= "protein modification and maintenance molecule"
XX
XX WO200246383-A2.
PN
XX 13-JUN-2002.
PD
XX 05-DEC-2001; 2001WO-US046964.
PF

XX 08-DEC-2000; 2000US-0254399P.
PR 21-DEC-2000; 2000US-0257803P.
PR 05-JAN-2001; 2001US-0260110P.
PR 19-JAN-2001; 2001US-0262851P.
PR 25-JAN-2001; 2001US-0264623P.
XX (INCY-) INCYTE GENOMICS INC.
XX Yue H, Azimzai Y, Kallick DA, Baughn MR, Griffin JA, Swarnakar A;
PI Lal PG, Wallia NK, Hafalia AJA, Gandhi AR, Au-Young J, Elliott VS;
PI Rangkumar J, Thangavelu K, Lu Y, Warren BA, Lu DAM, Lee EA;
PI Tribouley CM, Arvizu C, Delegeane AM, Yao MG, Khan FA;
PI Sanjanwala MM;
XX WPI; 2002-519664/55.
DR P-PSDB; ABB98135.
XX
PT New isolated Protein Modification and Maintenance polypeptides, useful
PT for diagnosis, and treatment of e.g. gastrointestinal disorders.
XX
PS Claim 5; Page 196; 200pp; English.
XX
CC The invention relates to an isolated Protein Modification and Maintenance
CC (PMM) polypeptide. Polypeptides of the invention may be used in the
CC diagnosis, treatment and prevention of disorders associated with
CC decreased expression or activity of PMM. These include gastrointestinal
CC disorders (e.g. Crohn's disease), cardiovascular disorders (e.g.
CC hypertension), autoimmune/inflammatory disorders (e.g. anaemia), cell
CC proliferative disorders, developmental disorders, epithelial disorders
CC (e.g. scabies), neurological disorders (e.g. Alzheimer's disease)
CC reproductive disorders (e.g. ectopic pregnancy), and in gene therapy or a
CC vaccine for such diseases. They may also be used in the assessment of the
CC effects of exogenous compound on the expression of nucleic acid and amino
CC acid sequences of protein modification and maintenance molecules. The
CC current sequence represents a human PMM encoding sequence of the
CC invention, encoding a polypeptide which has been found to have homology
CC with rat prostasin
XX
SQ Sequence 2681 BP; 431 A; 919 C; 861 G; 470 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.: 8,09e-240 Length: 2681
Score: 272.00 Matches: 272
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 76.19% Indels: 0
DB: Gaps: 0

US-10-037-417-46 (1-357) x ABQ75956 (1-2681)
Qy 38 CysGlyArgProGluProSerAlaArgIleValGlyGlySerAsnAlaGlnProGlyThr 57
Db 112 TGGGGCGCCCTGAGCCCTCGGCCCGCATCTGGGGGGCTCAAAACGCGCAGCGGGCACC 171
Qy 58 TrpProTrpGlnValSerLeuHisGlyGlyHisIleCysGlyGlySerLeu 77
Db 172 TGGCCTTGGCAAGTGAGCTGCACCATGGAGGTGGCCACATCTCGGGGGTCCCTCATC 231
Qy 78 AlaProSerTrpValLeuSerAlaAlaHisCysPheMetThrAsnGlyThrLeuGluPro 97
Db 232 GCCCCCTCTGGGTCCCTCTCCGCTGCTCAGTTCATGACGAATGGAGCGCTGGAGGCC 291
Qy 98 AlaAlaGluTrpSerValLeuLeuGlyValHisSerGlnAspGlyProLeuAspGlyAla 117
Db 292 GCGGCGGAGTGGTGGTACTCTGCGCGTGTCACCTCCAGAGACGGGCCCTCGACGGCGC 351
Qy 118 HisThrArgAlaValAlaAlaIleValProAlaAsnTyrSerGlnValGluLeuGly 137
Db 352 CACACCCGCGAGTGGCGGCCCATCTGTTGTCGCGGCCCACTACAGCAAGTGGAGCTGGGC 411
Qy 138 AlaAspLeuAlaLeuLeuArgLeuAlaSerProAlaSerLeuGlyProAlaValTrpPro 157

Db 412 GCGACCTGCGCTGCTGCGCTCACCAGCCTGGGCGCCGCGCTGTCGCT 471
 QY 158 ValCysLeuProArgAlaSerHisArgPheValHisGlyThrAlaCysTrpAlaThrGly 177
 Db 472 GTCTGCTGCGCGCTCAGCCGCTTGTGCACGGCACCGCTGCTGGGCGACCGGC 531
 QY 178 TrpGlyAspValGlnGluAlaAspProLeuProLeuProTrpValLeuGlnGluValGlu 197
 Db 532 TGGGGAGACGTCAGGAGGAGAGTCTCTGCTCTCCCTGCTGCTACAGGAAGTGAG 591
 QY 198 LeuArgLeuLeuGlyGluAlaThrCysGlnCysLeuTyrSerGlnProGlyProPheAsn 217
 Db 592 CTAGGCTGCTGGCGGAGCCACCTGCAATGCTCTACAGCCAGCGCGCTCCCTTCAAC 651
 QY 218 LeuThrLeuGlnIleLeuProGlyMetLeuCysAlaGlyTyrProGluGlyArgArgAsp 237
 Db 652 CTCACITCCAGATATTGCCAGGGATGCTGTGTGCTACCCAGAGGCGCGAGGGAC 711
 QY 238 ThrCysGlnGlyAspSerGlyGlyProLeuValCysGluGluGlyArgTrpPheGln 257
 Db 712 ACCTGCCAGGGTGACTCTGGGGGGCCCTGCTGTGTGAGGAAGCGCGCTGCTCCAG 771
 QY 258 AlaGlyIleThrSerPheGlyPheGlyCysGlyArgArgGlnArgGlnValPheThr 277
 Db 772 GCGAGAAATCACCAGCTTGTGCTGTGGCTGGAGAAACCGCCCTGGAGTTTCACT 831
 QY 278 AlaValAlaThrTyrGluAlaTrpIleArgGluGlnValMetGlySerGluProGlyPro 297
 Db 832 GCTGTGCTACCTATGAGGATGATGATACGGGAGCAGGTGATGGTTCAGAGCCTGGGCT 891
 QY 298 AlaPheProThrGlnProGlnLysThrGlnSerAsp 309
 Db 892 GCCTTTCCACCCAGCCCGCAGAGACCCAGTCAGAT 927
 RESULT 5
 AAK94500
 ID AAK94500 standard; cDNA; 2810 BP.
 XX AC AAK94500;
 XX 06-NOV-2001 (first entry)
 XX Human full-length cDNA, SEQ ID NO: 3346.
 XX Human; full length cDNA; cDNA synthesis; oligo-capping; ss.
 XX Homo sapiens.
 XX EF1130094-A2.
 XX 05-SEP-2001.
 XX 07-JUL-2000; 2000EP-00114089.
 XX 08-JUL-1999; 99JP-00194486.
 XX 11-JAN-2000; 2000JP-00118774.
 XX 02-MAY-2000; 2000JP-00183765.
 XX (HELI-) HELIX RES INST.
 XX Ota T, Nishikawa T, Isogai T, Hayashi K, Ishii S, Kawai Y;
 XX Wakamatsu A, Sugiyama T, Nagai K, Kojima S, Otsuki T, Koga H;
 XX WPI; 2001-524255/58.
 XX P-PSDB; AAM93568.
 XX 830 Primers useful for synthesizing full length cDNA clones and their use
 XX in genetic manipulation.
 XX Claim 8; SEQ ID NO 3346; 1380pp + Sequence Listing; English.
 XX The invention relates to primers for synthesizing full length cDNA
 XX clones. 830 cDNA molecules encoding a human protein have been isolated

CC and nucleotide sequences of 5'- and 3'-ends of the cDNA molecules have
 CC been determined. Primers for synthesizing the full length cDNA are useful
 CC for clarifying the function of the protein encoded by the cDNA. The full
 CC length clones were obtained by construction of full length enriched cDNA
 CC libraries that were synthesised by the oligo-capping method. The primers
 CC enable the production of the full length cDNA easily without any special
 CC methods. The present sequence is a full length human cDNA of the
 CC invention. Note: The sequence data for this patent did not form part of
 CC the printed specification, but was obtained in CD-ROM format directly
 CC from EPO
 XX
 SQ Sequence 2810 BP; 436 A; 964 C; 902 G; 508 T; 0 U; 0 Other;
 Alignment Scores:
 Pred. No.: 1-08e-229 Length: 2810
 Score: 261.00 Matches: 261
 Percent Similarity: 100.00% Conservatives: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 73.11% Indels: 0
 DB: 4 Gaps: 0
 US-10-037-417-46 (1-357) x.AAK94500 (1-2810)
 QY 49 GlyGlySerAsnAlaGlnProGlyThrTrpProTrpGlnValSerLeuHisHisGlyGly 69
 Db 177 GGGGGCTCAACCGCGCAGCGCGGCACTGGCTTGGCAAGTGAGCTGCACCATGGAGGT 236
 QY 69 GlyHisIleCysGlyGlySerLeuIleAlaProSerTrpValLeuSerAlaHisCys 88
 Db 237 GGCACATCTGGGGGGCTCCCTCATCGCCCTCTCTGGGCTCTCTCCGCTGCTACTGT 296
 QY 89 PheMetThrAsnGlyThrLeuGluProAlaAlaGluTrpSerValLeuLeuGlyValHis 108
 Db 297 TTCATGACGAATGGGACGCTGGAGCCCGCGCGAGTGTGCTACTCTGGCGGTGCAC 356
 QY 109 SerGlnAspGlyProLeuAspGlyAlaHisThrArgAlaValAlaAlaIleValPro 128
 Db 357 TCCAGGACGGGCGCTGGAGCGCGCACACCGCGCAGTGGCGCCCATCTGTGTGTCGG 416
 QY 129 AlaAsnTyrSerGlnValGluLeuGlyAlaAspLeuAlaLeuArgLeuAlaSerPro 148
 Db 417 GCAACATCAGCAAGTGGAGCTGGGCGCGCACTGGCCCTCTGTGGCTGGCCCTCACCC 476
 QY 149 AlaSerLeuGlyProAlaValTrpProValCysLeuProArgAlaSerHisArgPheVal 168
 Db 477 GCGAGCTGGCGCGCGCTGTGGCTGTCTGCTGCCCGCGCCTCACACCGCTTCGTG 536
 QY 169 HisGlyThrAlaCysTrpAlaThrGlyTrpGlyAspValGlnGluAlaAspProLeuPro 188
 Db 537 CACGGCACCGCTGTGGGCCACCGCTGGGAGACGTCACAGAGGACAGATCTCTGCT 596
 QY 189 LeuProTrpValLeuGlnGluValGluLeuArgLeuLeuGlyGluAlaThrCysGlnCys 208
 Db 597 CTCCTCTGGTGTACAGGAAGTGGAGCTAAGGCTGTGGGCGAGGCCACTGTCTCATGT 556
 QY 209 LeuTyrSerGlnProGlyProPheAsnLeuThrLeuGlnIleLeuProGlyMetLeuCys 228
 Db 657 CTCACAGCCAGCGCGCTCCCTTCAACCTCACTCTCCAGATATTGCCAGGATGTGTGT 716
 QY 229 AlaGlyTyrProGluGlyArgArgAspThrCysGlnGlyAspSerClyGlyProLeuVal 248
 Db 717 GTTGGCTACCCAGAGGGCGGAGCACCTGCCAGGCGAGTCTTGGGGGGCGCCCTGGTC 776
 QY 249 CysGluGluGlyGlyArgTrpPheGlnAlaGlyIleThrSerPheGlyPheGlyCysGly 268
 Db 777 TGTGAGGAAGCGCGCTGTGTTCAGGAGCAAGATCACAGGCTTTGGCTTTGGCTTGTGA 836
 QY 269 ArgArgAsnArgProGlyValPheThrAlaValAlaThrTyrGluAlaTrpIleArgGlu 288
 Db 837 CGGAGAAACCGCCCTGGAGTTTTCATCTGTGTGGCTACCTATGAGCATGGATACGGAG 896
 QY 289 GlnValMetGlySerGluProGlyProAlaPheProThrGlnProGlnLysThrGlnSer 308


```

XX WO200175067-A2.
PN
XX
XX
XX 11-OCT-2001.
XX
XX 30-MAR-2001; 2001WO-US008631.
XX
XX 31-MAR-2000; 2000US-00540217.
PR
XX 23-AUG-2000; 2000US-00649167.
PR
XX (HYSE-) HYSEQ INC.
PA
XX
XX Drmanac RT, Liu C, Tang YT;
PI
XX
XX WPI; 2001-639362/73.
DR
XX P-PSDB; ABG04853.
DR
XX
XX New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity.
PT
XX
XX Claim 1; SEQ ID NO 4844; 103pp; English.
PS
XX
XX The invention relates to isolated polynucleotide (I) and polypeptide (II)
CC sequences. (I) is useful as hybridisation probes, polymerase chain
CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
CC and in recombinant production of (II). The polynucleotides are also used
CC in diagnostics as expressed sequence tags for identifying expressed
CC genes. (I) is useful in gene therapy techniques to restore normal
CC activity of (II) or to treat disease states involving (II). (II) is
CC useful for generating antibodies against it, detecting or quantitating a
CC polypeptide in tissue, as molecular weight markers and as a food
CC supplement. (II) and its binding partners are useful in medical imaging
CC of sites expressing (II). (I) and (II) are useful for treating disorders
CC involving aberrant protein expression or biological activities. The
CC polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. AAS64197-AAS94564 represent novel human diagnostic
CC coding sequences of the invention. Note: The sequence data for this
CC patent did not appear in the printed specification, but was obtained in
CC electronic format directly from WIPO at
XX ftp.wipo.int/pub/published_pct_sequences
XX
XX Sequence 597 BP; 97 A; 217 C; 196 G; 87 T; 0 U; 0 Other;
SQ

```

Alignment Scores:

Pred. No.:	Length:	Score:
9,06e-112	597	132.00
Matches:	132	
Conservative:	0	
Mismatches:	0	
Indels:	0	
Gaps:	5	

US-10-037-417-46 (1-357) x AAS69040 (1-597)

```

Qy 109 SerGlnAspGlyProLeuAspGlyAlaHisThrArgAlaValAlaAlaValValPro 128
Db 52 TCCGAGGACGGGGCCCTGGACGGCGCGCACACCCGCGAGTGGCGGCATCGTGGTCCG 111
Qy 129 AlaSerLeuGlnValGluLeuGlyAlaAspLeuAlaLeuLeuAlaSerPro 148
Db 112 GCCAACTACACCAAGTGGAGCTGGGCGCCGACCTGGCCCTGCTGGCGCTGCACCC 171
Qy 149 AlaSerLeuGlyProAlaValTrpProValCysLeuProArgAlaSerHisArgPheVal 168
Db 172 GCCAGCCTGGGCGCGCGCTGTGGCTGTCTGCTGTGGCGCGGCTGCACACCGTTCGTG 231
Qy 169 HisGlyThrAlaCysTrpAlaThrClyTrpGlyAspValGlnGluAlaAspProLeuPro 188
Db 232 CACGGCACCGCTGTCTGGGCGCACCGGCTGGGAGACGTCACGAGGCGAGATCCTCTGCCT 291

```

Alignment Scores:

Pred. No.:	Length:	Score:
1e-111	670	132.00
Matches:	132	
Conservative:	0	
Mismatches:	0	
Indels:	0	
Gaps:	4	

US-10-037-417-46 (1-357) x AAK92190 (1-670)

```

Qy 189 leuProTrpValLeuGlnGluValGluLeuArgLeuLeuGlyGluAlaThrCysGlnCys 208
Db 292 CTCCCTGGGTGCTACAGGAAGTGGAGTAAAGGTGCTGGGCGAGGCCACCTGTCAATGT 351
Qy 209 LeuTrpSerGlnProGlyProPheAsnLeuThrLeuGlnTrpLeuProGlyMetLeuCys 228
Db 352 CTCACAGCCAGCCCGGTCCTTCAACCTACTCTCCAGATATTGCCAGGAGTGTGTGT 411
Qy 229 AlaGlyTrpProGluGlyArgArgAspThrCysGln 240
Db 412 GCTGGCTACCCAGAGGGCGCGGACACACCTGGCCAG 447

```

RESULT 8

AAK92190

ID AAK92190 standard; cDNA; 670 BP.

XX

AC AAK92190;

XX

DT 06-NOV-2001 (first entry)

XX

DE Human cDNA 5'-end sequence, SEQ ID NO: 650.

XX

XX Human; full length cDNA; cDNA synthesis; oligo-capping; ss.

XX

XX Homo sapiens.

OS

XX

PN EP1130094-A2.

XX

PD 05-SEP-2001.

XX

PF 07-JUL-2000; 2000EP-00114089.

XX

PR 08-JUL-1999; 95JP-00194486.

XX

PR 11-JAN-2000; 2000JP-00118774.

XX

PR 02-MAY-2000; 2000JP-00183765.

XX

XX (HELI-) HELIX RES INST.

PA

XX

XX Ota T, Nishikawa T, Isogai T, Hayashi K, Ishii S, Kawai Y;

PI Wakamatsu A, Sugiyama T, Nagai K, Kojima S, Otsuki T, Koga H;

XX

XX WPI; 2001-524255/58.

XX

XX 830 Primers useful for synthesizing full length cDNA clones and their use

PT in genetic manipulation.

PS

PS Claim 2; SEQ ID NO 650; 1380pp + Sequence Listing; English.

XX

XX The invention relates to primers for synthesising full length cDNA

CC clones. 830 cDNA molecules encoding a human protein have been isolated

CC and nucleotide sequences of 5'- and 3'-ends of the cDNA molecules have

CC been determined. Primers for synthesising the full length cDNA are useful

CC for clarifying the function of the protein encoded by the cDNA. The full

CC length clones were obtained by construction of full length enriched cDNA

CC libraries that were synthesised by the oligo-capping method. The primers

CC enable the production of the full length cDNA easily without any special

CC methods. The present sequence is the nucleotide sequence of the 5'-end of

CC a cDNA provided in the invention. Note: The sequence data for this patent

CC did not form part of the printed specification, but was obtained in CD-

CC ROM format directly from EPO

XX

SQ Sequence 670 BP; 97 A; 245 C; 205 G; 118 T; 0 U; 5 Other;

XX 23-JUN-2000; 2000US-0213588P.
PR 20-MAR-2001; 2001US-0276909P.
XX (FARB) BAYER AG.
FA Xiao Y, Morozov V;
XX WPI; 2002-114576/15.
XX
XX Novel human prostatic-like serine protease polypeptide and polynucleotide
PT which can be regulated for treating metastasis of malignant cells,
PT inflammation, atherosclerosis, neurodegenerative disease and infections.
XX
XX Disclosure; Fig 3; 11pp; English.
XX
XX This invention comprises the cDNA and protein sequences of an isolated
CC prostatic-like serine protease and reagents and methods for regulating
CC the human prostatic-like enzyme activity. Prostatic is a trypsin-like
CC serine protease purified from human seminal fluid. An antibody specific
CC for prostatic-like serine protease is useful for immunodetection and
CC diagnosis of micro-metastases, autoimmune lesions and renal failure in
CC biopsy specimens, plasma samples and body fluids. The antibody may be
CC used to modulate enzyme activity in a disease, such as metastasis of
CC malignant cells, tumour angiogenesis, inflammation, chronic obstructive
CC pulmonary disease (COPD), atherosclerosis, neurodegenerative disease
CC (e.g. prion protein amyloid plaques of Gensmann-Straussler Syndrome,
CC Creutzfeldt-Jakob disease, Scrapie) or infection, particularly viral
CC infection. The human prostatic-like serine protease gene provides a
CC therapeutic target of decreasing human prostatic-like serine protease
CC activity, in particular for treating or preventing metastatic cancer. The
CC agonists and antagonists of the nucleotide sequence may be used to mimic,
CC augment and inhibit the enzyme activity which may be useful to treat
CC osteoporosis, Paget's disease and degradation of bone implants
CC particularly dental implants. Altered levels of human prostatic-like
CC serine protease activity inhibits both smooth muscle cell proliferation
CC and lipid accumulation and inhibit the progression of restenosis and
CC atherosclerosis. The nucleic acid sequence is also useful in diagnostic
CC assays for detecting diseases and abnormalities or susceptibility to
CC diseases related to the presence of mutations in nucleic acid sequences
CC which encode the enzyme. The present sequence represents the human
CC prostatic-like serine protease #2 nucleotide sequence of the invention
XX
SQ Sequence 456 BP; 74 A; 151 C; 167 G; 62 T; 0 U; 2 Other;

Alignment Scores:
Pred. No.: 3,37e-108 Length: 456
Score: 128.00 Matches: 128
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 35.85% Indels: 0
DB: 6 Gaps: 0

US-10-037-417-46 (1-357) x ABK13566 (1-456)

QY 121 AlaValAlaAlaValValProAlaAsnTyrSerGlnValGluLeuGlyAlaAspLeu 140
Db 384 CGAGTGGCGCCATCGTGGTGGCGGCACTACAGCAAGTGGAGTGGCGCGCACCTG 325
QY 141 AlaLeuLeuArgLeuAlaSerProAlaSerLeuGlyProAlaValTrpProValCysLeu 160
Db 324 GCCTCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTG 265
QY 161 ProArgAlaSerHisArgPheValHisGlyThrAlaCysTrpAlaThrGlyTrpGlyAsp 180
Db 264 CCCCGGCTTCACACCGCTTCGTCGACGCGACCGCTGGCGCCACCGCTGGCGAGAC 205
QY 181 ValGlnGluAlaAspProLeuProLeuProTrpValGlnGluValGlnLeuArgLeu 200
Db 204 GTCCAGGAGGAGATCTCTGCTCTCTCCCTGGGTGCTACAGGAAGTGGAGCTGAAGGCTG 145
QY 201 LeuGlyGluAlaThrCysGlnCysLeuTyrSerGlnProGlyProPheAsnLeuThrLeu 220

Db 144 CTGGCGGAGGCCACCTGTCAATGTCTTACAGCAGCCCGGTCCCTTCAACCTCACTCTC 85
QY 221 GlnLeuLeuProGlyMetLeuCysAlaGlyTyrProGluGlyArgArgAspThrCysGln 240
Db 84 CAGATATTGCGAGGGATGCTGTGTGGTGTACCCAGAGGGCGCGAGACACCTGCCAG 25
QY 241 GlyAspSerGlyGlyProLeuVal 248
Db 24 GGTGACTCTGGGGGCGCCCTGGTC 1
RESULT 11
ID AAS71079 standard; cDNA; 2298 BP.
XX AC AAS71079;
XX DT 13-FEB-2002 (first entry)
XX DE DNA encoding novel human diagnostic protein #6883.
XX KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
XX food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX OS Homo sapiens.
XX PN WO200175067-A2.
XX PD 11-OCT-2001.
XX PF 30-MAR-2001; 2001WO-US008631.
XX PR 31-MAR-2000; 2000US-00540217.
XX PR 23-AUG-2000; 2000US-00649167.
XX PA (HYSE-) HYSEQ INC.
XX PI Drmanac RT, Liu C, Tang YT;
XX WPI; 2001-639362/73.
XX P-PSDB; ASB06892.
XX New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity.
XX Claim 1; SEQ ID NO 6883; 103pp; English.
XX The invention relates to isolated polynucleotide (I) and polypeptide (II)
CC sequences. (I) is useful as hybridisation probes, polymerase chain
CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
CC and in recombinant production of (II). The polynucleotides are also used
CC in diagnostics as expressed sequence tags for identifying expressed
CC genes. (I) is useful in gene therapy techniques to restore normal
CC activity of (II) or to treat disease states involving (II). (II) is
CC useful for generating antibodies against it, detecting or quantitating a
CC polypeptide in tissue, as molecular weight markers and as a food
CC supplement. (II) and its binding partners are useful in medical imaging
CC of sites expressing (II). (I) and (II) are useful for treating disorders
CC involving aberrant protein expression or biological activity. The
CC polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. AAS64197-AAS94564 represent novel human diagnostic
CC coding sequences of the invention. Note: The sequence data for this
CC patent did not appear in the printed specification, but was obtained in
CC electronic format directly from WIPO at
XX ftp.wipo.int/pub/published_pct_sequences
SQ Sequence 2298 BP; 468 A; 774 C; 706 G; 350 T; 0 U; 0 Other;

Alignment Scores:


```
Pred. No.: 9.73e-38 Length: 2298
Score: 52.00 Matches: 52
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 14.57% Indels: 0
DB: 5 Gaps: 0

US-10-037-417-46 (1-357) x AAS71079 (1-2298)

QY 258 AlaGlyIleThrSerPheGlyCysGlyArgArgAsnArgProGlyValPheThr 277
Db 424 GCAGAAATCACAGCTTGGCTTTGGCTGTGGAGGAGAAACCGCCCGGAGTTTCACT 483
QY 278 AlaValAlaThrTyrGluAlaTrpIleArgGluInValMetGlySerGluProGlyPro 297
Db 484 GCTGTGGCTACTATGAGCATGTGATACGGGAGCAGGTGATGGTTTCAGAGCCTGGGCT 543
QY 298 AlaPheProThrGlnProGlnLysThrGlnSerAsp 309
Db 544 GCCTTTCCACCCAGCCGCCAGAGACCCAGTCAGAT 579

RESULT 12
AAH78206
ID AAH78206 standard; DNA; 768 BP.
XX AC
AC AAH78206;
XX AC
XX 26-NOV-2001 (first entry)
XX DE
DE Nucleotide sequence of a human secreted polypeptide.
XX KW Human; secreted polypeptide; nervous disease; muscular disease; tumour;
KW Gastrointestinal ulceration; spinal cord disease; trachea disease;
KW thyroid gland disease; ovary disease; prostate disease; heart disease;
KW renal gland disease; small intestine disease; thymus disease;
KW lymph node disease; muscular system disease; colon disease;
KW lipase deficiency; cystic fibrosis; pancreatitis; clot formation;
KW myocardial infarction; angioplasty; liver disease; coagulation disorder;
KW microbial disease; immune disorder; inflammation; transplant rejection;
KW bone thickness; bone density; ferroxidase loss; apoptosis;
KW vascular smooth cell proliferation; vaccine; ss.
XX OS
OS Homo sapiens.
XX FH
XX Key Location/Qualifiers
XX CDS 1..768
XX FT /*tag= a
XX FT /product= "secreted polypeptide"
XX FT
XX PN WO200166690-A2.
XX PD
XX PD 13-SEP-2001.
XX XX
XX XX 05-MAR-2001; 2001WO-US007143.
XX PF
XX XX 06-MAR-2000; 2000US-0187107P.
XX PR 13-MAR-2000; 2000US-0188916P.
XX PR 03-OCT-2000; 2000US-0236874P.
XX PR 03-OCT-2000; 2000US-0237846P.
XX XX
XX (SMIK ) SMITHKLINE BEECHAM CORP.
XX PA (SMIK ) SMITHKLINE BEECHAM PLC.
XX XX
XX Agarwal P, Murdoch PR, Rizvi SK, Smith RF, Xiang Z;
XX PI
XX WPI: 2001-570768/64.
XX DR P-PSDB; AAG67514.
XX XX
XX Novel isolated secreted polypeptide useful for treating nervous and
XX PT muscular diseases, gastrointestinal ulceration, coagulation and immune
XX PT disorders, microbial diseases, inflammation and transplant rejection.
XX PT
XX Claim 2; Page 44-45; 102pp; English.
XX PS
```

The present sequence encodes a human secreted polypeptide. The secreted polypeptides and polynucleotides are useful for treating nervous and muscular diseases, for inhibiting tumour formation and metastasis, for treating gastrointestinal ulceration, for preventing and treating diseases in spinal cord, thyroid gland, ovary, prostate, renal gland, small intestine, heart, thymus, lymph node, muscular system and colon, for treating lipase deficiency in cystic fibrosis and pancreatitis, for treating undesirable clot formation such as myocardial infarction, during angioplasty and all surgical procedures that require decreased blood clot formation, for treating liver diseases, coagulation disorders and microbial diseases, for treating immune disorders, for treating inflammation and transplant rejection, for enhancing bone thickness and increasing bone density, for reducing the loss of essential ferroxidases, for suppressing apoptosis, and for regulating vascular smooth cell proliferation. They may also be used as vaccines

Sequence 768 BP; 125 A; 280 C; 224 G; 139 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.:	9.83e-32	Length:	768
Score:	45.00	Matches:	45
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	12.61%	Indels:	0
DB:	4	Gaps:	0

US-10-037-417-46 (1-357) x AAH78206 (1-768)

QY 47 IleValGlyGlySerAsnAlaGlnProGlyThrTrpProTrpGlnValSerLeuHisHis 66
Db 1 ATCTGGGGGGCTCAAAACGGCGACGCGGCTGGGCTTGGCAAGTGAGCTGCACCAT 60
QY 67 GlyGlyGlyHisIleCysGlyGlySerLeuIleAlaProSerTrpValLeuSerAlaAla 86
Db 61 GGAGGTGGCCACATCTCGGGGGCTCCCTCATGCCCCCTCTGGGCTCTCCGCTCT 120
QY 87 HisCysPheMetThr 91
Db 121 CACTGTTTCATGAGG 135

RESULT 13
ABA21236/c
ID ABA21236 standard; DNA; 8899 BP.
XX AC
AC ABA21236;
XX AC
XX 23-JAN-2002 (first entry)
XX DE
DE Human nervous system related polynucleotide SEQ ID NO 13567.
XX KW Human; nootropic; neuroprotective; cytostatic; dermatological; virucide;
KW immunosuppressive; antiinflammatory; anti-HIV; antibacterial; vulnery;
KW antiparkinsonian; antisickling; antianaemic; antiarthritic; cancer;
KW antineumatic; hepatotropic; cerebroprotective; antiinflammatory;
KW antiallergic; antidiabetic; antidiabetic; antidiabetic; antifungal;
KW antiparasitic; cardiac; immune disorder; cardiovascular disorder;
KW neurological disease; infection; nephrotropic; gene therapy; vaccine; ds.
XX OS
OS Homo sapiens.
XX PN WO200159063-A2.
XX XX
XX PD 16-AUG-2001.
XX XX
XX PD 17-JAN-2001; 2001WO-US001334.
XX PF
XX 31-JAN-2000; 2000US-0179065P.
XX PR 04-FEB-2000; 2000US-0180628P.
XX PR 24-FEB-2000; 2000US-0184664P.
XX PR 02-MAR-2000; 2000US-0186350P.
XX PR 16-MAR-2000; 2000US-0189874P.
XX PR 17-MAR-2000; 2000US-0190076P.

us-10-037-417-46.olip2n.rng

Thu Feb 26 13:50:24 2004

PR	18-APR-2000;	2000US-0198123P.	PR	20-OCT-2000;	2000US-0240960P.
PR	19-MAY-2000;	2000US-0205515P.	PR	20-OCT-2000;	2000US-0241785P.
PR	07-JUN-2000;	2000US-0209467P.	PR	20-OCT-2000;	2000US-0241786P.
PR	28-JUN-2000;	2000US-0214896P.	PR	20-OCT-2000;	2000US-0241787P.
PR	30-JUN-2000;	2000US-0215135P.	PR	20-OCT-2000;	2000US-0241808P.
PR	07-JUL-2000;	2000US-0216647P.	PR	20-OCT-2000;	2000US-0241809P.
PR	07-JUL-2000;	2000US-0216880P.	PR	20-OCT-2000;	2000US-0241826P.
PR	11-JUL-2000;	2000US-0217487P.	PR	20-OCT-2000;	2000US-0242221P.
PR	11-JUL-2000;	2000US-0217496P.	PR	01-NOV-2000;	2000US-0244617P.
PR	14-JUL-2000;	2000US-0218290P.	PR	08-NOV-2000;	2000US-0246474P.
PR	26-JUL-2000;	2000US-0220963P.	PR	08-NOV-2000;	2000US-0246475P.
PR	26-JUL-2000;	2000US-0220964P.	PR	08-NOV-2000;	2000US-0246476P.
PR	14-AUG-2000;	2000US-0224518P.	PR	08-NOV-2000;	2000US-0246477P.
PR	14-AUG-2000;	2000US-0224519P.	PR	08-NOV-2000;	2000US-0246478P.
PR	14-AUG-2000;	2000US-0225213P.	PR	08-NOV-2000;	2000US-0246523P.
PR	14-AUG-2000;	2000US-0225214P.	PR	08-NOV-2000;	2000US-0246524P.
PR	14-AUG-2000;	2000US-0225266P.	PR	08-NOV-2000;	2000US-0246525P.
PR	14-AUG-2000;	2000US-0225267P.	PR	08-NOV-2000;	2000US-0246526P.
PR	14-AUG-2000;	2000US-0225268P.	PR	08-NOV-2000;	2000US-0246527P.
PR	14-AUG-2000;	2000US-0225270P.	PR	08-NOV-2000;	2000US-0246528P.
PR	14-AUG-2000;	2000US-0225447P.	PR	08-NOV-2000;	2000US-0246532P.
PR	14-AUG-2000;	2000US-0225757P.	PR	08-NOV-2000;	2000US-0246610P.
PR	14-AUG-2000;	2000US-0225758P.	PR	08-NOV-2000;	2000US-0246611P.
PR	14-AUG-2000;	2000US-0225759P.	PR	08-NOV-2000;	2000US-0246613P.
PR	14-AUG-2000;	2000US-0226799P.	PR	08-NOV-2000;	2000US-0249207P.
PR	22-AUG-2000;	2000US-0226681P.	PR	17-NOV-2000;	2000US-0249208P.
PR	22-AUG-2000;	2000US-0226686P.	PR	17-NOV-2000;	2000US-0249209P.
PR	22-AUG-2000;	2000US-0227182P.	PR	17-NOV-2000;	2000US-0249210P.
PR	23-AUG-2000;	2000US-0227009P.	PR	17-NOV-2000;	2000US-0249211P.
PR	30-AUG-2000;	2000US-0228924P.	PR	17-NOV-2000;	2000US-0249212P.
PR	01-SEP-2000;	2000US-0229343P.	PR	17-NOV-2000;	2000US-0249213P.
PR	01-SEP-2000;	2000US-0229344P.	PR	17-NOV-2000;	2000US-0249214P.
PR	01-SEP-2000;	2000US-0229345P.	PR	17-NOV-2000;	2000US-0249216P.
PR	03-SEP-2000;	2000US-0229509P.	PR	17-NOV-2000;	2000US-0249217P.
PR	05-SEP-2000;	2000US-0229513P.	PR	17-NOV-2000;	2000US-0249218P.
PR	06-SEP-2000;	2000US-0230437P.	PR	17-NOV-2000;	2000US-0249244P.
PR	06-SEP-2000;	2000US-0230438P.	PR	17-NOV-2000;	2000US-0249245P.
PR	08-SEP-2000;	2000US-0231242P.	PR	17-NOV-2000;	2000US-0249264P.
PR	08-SEP-2000;	2000US-0231243P.	PR	17-NOV-2000;	2000US-0249265P.
PR	08-SEP-2000;	2000US-0231244P.	PR	17-NOV-2000;	2000US-0249297P.
PR	08-SEP-2000;	2000US-0231413P.	PR	17-NOV-2000;	2000US-0249299P.
PR	08-SEP-2000;	2000US-0231414P.	PR	17-NOV-2000;	2000US-0249300P.
PR	08-SEP-2000;	2000US-0232080P.	PR	01-DEC-2000;	2000US-0250391P.
PR	08-SEP-2000;	2000US-0232081P.	PR	01-DEC-2000;	2000US-0251160P.
PR	12-SEP-2000;	2000US-0231968P.	PR	05-DEC-2000;	2000US-0251030P.
PR	14-SEP-2000;	2000US-0232397P.	PR	05-DEC-2000;	2000US-0251988P.
PR	14-SEP-2000;	2000US-0232398P.	PR	05-DEC-2000;	2000US-0256719P.
PR	14-SEP-2000;	2000US-0232399P.	PR	06-DEC-2000;	2000US-0251479P.
PR	14-SEP-2000;	2000US-0232400P.	PR	08-DEC-2000;	2000US-0251856P.
PR	14-SEP-2000;	2000US-0232401P.	PR	08-DEC-2000;	2000US-0251866P.
PR	14-SEP-2000;	2000US-0233063P.	PR	08-DEC-2000;	2000US-0251869P.
PR	14-SEP-2000;	2000US-0233064P.	PR	08-DEC-2000;	2000US-0251989P.
PR	14-SEP-2000;	2000US-0233065P.	PR	08-DEC-2000;	2000US-0251990P.
PR	21-SEP-2000;	2000US-0234223P.	PR	11-DEC-2000;	2000US-0254097P.
PR	21-SEP-2000;	2000US-0234274P.	PR	05-JAN-2001;	2001US-0259678P.
PR	25-SEP-2000;	2000US-0234997P.	XX		
PR	25-SEP-2000;	2000US-0234998P.	PA	(HUMA-) HUMAN GENOME SCI INC.	
PR	26-SEP-2000;	2000US-0235484P.	XX		
PR	27-SEP-2000;	2000US-0235834P.	PI	Rosen CA, Barash SC, Ruben SM;	
PR	27-SEP-2000;	2000US-0235836P.	XX		
PR	29-SEP-2000;	2000US-0236327P.	DR	WPI; 2001-541565/60.	
PR	29-SEP-2000;	2000US-0236367P.	XX		
PR	29-SEP-2000;	2000US-0236368P.	PT	Nucleic acids encoding 3224 human nervous system antigen polypeptides,	
PR	29-SEP-2000;	2000US-0236369P.	PT	useful for preventing, diagnosing and/or treating nervous system cancers	
PR	29-SEP-2000;	2000US-0236370P.	PT	and metastases.	
PR	02-OCT-2000;	2000US-0236802P.	XX	Disclosure; SEQ ID NO 13567; 1701pp + Sequence Listing; English.	
PR	02-OCT-2000;	2000US-0237037P.	PS		
PR	02-OCT-2000;	2000US-0237038P.	XX		
PR	02-OCT-2000;	2000US-0237039P.	XX		
PR	02-OCT-2000;	2000US-0237040P.	CC	The invention relates to novel genes (AB11004-ABA21534) and proteins	
PR	13-OCT-2000;	2000US-0239935P.	CC	(AB14678-AB18001) useful for preventing, treating or ameliorating	
PR	13-OCT-2000;	2000US-0239937P.	CC	medical conditions e.g. by protein or gene therapy. The genes are	

isolated from a range of human tissues disclosed in the specification. The nucleic acids, proteins, antibodies and (ant)agonists are useful in the diagnosis, treatment and prevention of: (a) cancer, e.g. breast and ovarian cancer and other cancers of the adrenal gland, bone, bone marrow, breast, gastrointestinal tract, liver, lung, or urogenital; (b) immune disorders e.g. Addison's disease, allergies, autoimmune haemolytic anaemia, autoimmune rheumatitis, diabetes mellitus, Crohn's disease, multiple sclerosis, rheumatoid arthritis and ulcerative colitis; (c) cardiovascular disorders such as myocardial ischaemia; (d) wound healing; (e) neurological diseases e.g. cerebral anoxia and epilepsy; and (f) infectious diseases such as viral, bacterial, fungal and parasitic infections. Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pat_sequences

Alignment Scores:	
Pred. No.:	4.73e-05
Score:	17.00
Percent Similarity:	100.00%
Best Local Similarity:	100.00%
Query Match:	4.76%
DB:	5
Gaps:	0
Indels:	0
Mismatches:	0
Conservative:	0
Matches:	17
Length:	8899

US-10-037-417-46 (1-357) x ABA21236 (1-8899)

Qy 341 LeuValLeuProTrpLeuSerProHisSerLeuLeuGlyLeuTrpGlyPhe 357
 | | | | | | | |
Db 747 CTTGTCTCTCCCTGGCTCTCTCCACACTCTCTCCTGGGCCTCTGGSGGTTTT 697

RESULT 14

ABN85393
ID ABN85393 standard; DNA: 1161 BP.

AC ABN85393:

21-OCT-2002 (first entry)

Human NOV14b, prostaticin-like protein, coding sequence.

Human: NOV14b; cytostatic; Cardiant; Antiinflammatory; Immunosuppressive;
 Anti-allergic; Nephrostatic; Anti-HIV; Anti-diabetic; Anorectic;
 Antiasthmatic; Nephrotropic; Hepatotropic; Neuroprotective; Nootropic;
 Antibacterial; Virucide; Antiparasitic; Relaxant; Anticonvulsant;
 Gene Therapy; NOV; cancer; heart disease; inflammation;
 autoimmune disorder; allergy; blood disorder; AIDS; diabetes; obesity;
 asthma; IGA nephropathy; cirrhosis; arthritis; Alzheimer's disease;
 infection; stroke; muscular dystrophy; epilepsy; wasting disorder;
 prostatic-like protein; chromosome 16; gene; ds.

OS Homo sapiens.

	Key	Location/Qualifiers
FT	CDS	1..1161
FT		/*tag= a
FT		/product= "NOV14b"

PN WO200255704-A2.

18-JUL-2002.

09-JAN-2002; 2002WO-US0000554.

PR 09-JAN-2001; 2001US-0260417P.

28-FEB-2001; 2001US-0272338P.

PR 18-APR-2001; 2001US-0284704P.

PA (CURA-) CURAGEN CORP.

PI Padigaru M, Li L, Zerhusen BD, Casman SJ, Shenov S. Spvtek KA:

PI	Zhong M, Gangolli EA, Burgess CE, Patturajan M, Vernet CAM;
PI	Taylor S, Tchernov VT, Miller CE, Guo X, Boldog FU, Grosse WM;
PI	Alsbrook JP, Gerlach V, Edinger S, Rothenberg ME, Ellerman K;
PI	Maddougall J, Malyankar U, Millet I, Peyman J, Smithson G;
PI	Gunther E, Stone DJ;
XX	
DR	WFI: 2002-590674/63.
DR	P-PSDB; AB98416.
XX	
PT	NOVX polypeptides and encoding polynucleotides, useful for preventing or
PT	treating NOVX-associated disorders e.g. cancer, inflammation, or
PT	Alzheimer's disease, and in chromosome mapping, tissue typing or
PT	pharmacogenomics.
XX	
PS	Claim 9; Page 99; 358pp; English.

The present sequence is a coding sequence for a NOV protein. The NOV proteins and coding sequences are useful for treating or preventing NOV-associated disorders or in the manufacture of a medicament for treating the disorders, such as cancer, heart disease, inflammation, autoimmune disorders, allergies, blood disorders, AIDS, diabetes, obesity, asthma, IgA nephropathy, cirrhosis, arthritis, Alzheimer's disease, infections (e.g. bacterial, viral, parasitic), stroke, muscular dystrophy, epilepsy and other wasting disorders associated with chronic diseases. NOV14b is a prostaticin-like protein, and the NOV14b coding sequence localises to chromosome 16.

Sequence 1161 BP; 219 A; 349 C; 364 G; 229 T; 0 U; 0 Other;

Alignment Scores:		
Pred. No.:	6.3e-05	Length:
Score:	16.00	Matches:
Percent Similarity:	100.00%	Conservative:
Best Local Similarity:	100.00%	Mismatches:
Query Match:	4.48%	Indels:
DB:	6	Gaps:

US-10-037-417-46 (1-357) x ABN85393 (1-1161)

Qy 1 MetAlaGlnLysGlyValLeuGlyProGlyGlnLeuGlyAlaValAla 16

Db 1 ATGCCCCAGAGGGGGTCTTGGGGCCCTGGCAGCTGGGGGCTGTGCC 48

RESULT 15

ID ABN85392 standard: DNA: 1726 BP.

AC ABN85392;

DT 21-OCT-2002 (first entry)

Human NOV14a, prostaticin-like protein, coding sequence.

Human; NOV14a; cytostatic; Cardiant; Antiinflammatory; Immunosuppressive;
 KW KW
 Anti-allergic; Haemostatic; Anti-HIV; Antidiabetic; Anorectic;
 KW KW
 Antiasthmatic; Nephrotoxic; Hepatotoxic; Neuroprotective; Nootropic;
 KW KW
 Antibacterial; Virucide; Antiparasitic; Relaxant; Anticonvulsant;
 KW KW
 Gene Therapy; NOV; cancer; heart disease; inflammation;
 KW KW
 autoimmune disorder; allergy; blood disorder; AIDS; diabetes; obesity;
 KW KW
 asthma; IGA nephropathy; cirrhosis; arthritis; Alzheimer's disease;
 KW KW
 infection; stroke; muscular dystrophy; epilepsy; wasting disorder;
 KW KW
 prostatic-like protein; chromosome 16p11.2; gene; ds.
 KW KW

OS Homo sapiens.

	Key	Location/Qualifiers
PH	CDS	229. .1152
FT		/*tag= a
FT		/product= "NOV14a"

PN WO200255704-A2.

PD 18-JUL-2002.

Thu Feb 26 13:50:24 2004

us-10-037-417-46.olip2n.rng

```
XX 09-JAN-2002; 2002WO-US000554.
XX
XX 09-JAN-2001; 2001US-0260417P.
XX
PR 10-JAN-2001; 2001US-0260831P.
PR
PR 28-FEB-2001; 2001US-0272338P.
PR
PR 09-MAR-2001; 2001US-0274876P.
PR
XX 18-APR-2001; 2001US-0284704P.
XX
XX (CURA-) CURAGEN CORP.
XX
XX Padigar M, Li L, Zerhusen BD, Casman SJ, Shenoy S, Spytek KA;
PI Zhong M, Gangalli EA, Burgess CE, Patturajan M, Vernet CM;
PI Taylor S, Tcherven V, Miller CE, Guo X, Boldog FL, Grose WM;
PI Alsobrook JP, Gerlach V, Edinger S, Rothenberg ME, Ellerman K;
PI Macdougall J, Malyankar U, Millet I, Feyman J, Smithson G;
PI Gunther E, Stone DJ;
XX
XX WPI; 2002-590674/63.
XX
XX P-PSDB; ABB98415.
XX
XX NOVX polypeptides and encoding polynucleotides, useful for preventing or
PT treating NOVX-associated disorders e.g. cancer, inflammation, or
PT Alzheimer's disease, and in chromosome mapping, tissue typing or
PT pharmacogenomics.
XX
XX Claim 9; Page 97-98; 358pp; English.
XX
XX The present sequence is a coding sequence for a NOV protein. The NOV
CC proteins and coding sequences are useful for treating or preventing NOV-
CC associated disorders or in the manufacture of a medicament for treating
CC the disorders, such as cancer, heart disease, inflammation, autoimmune
CC disorders, allergies, blood disorders, AIDS, diabetes, obesity, asthma,
CC IGA nephropathy, cirrhosis, arthritis, Alzheimer's disease, infections
CC (e.g. bacterial, viral, parasitic), stroke, muscular dystrophy, epilepsy,
CC and other wasting disorders associated with chronic diseases. NOV14a is a
CC prostaticin-like protein, and the NOV14a coding sequence localises to
CC chromosome 16p11.2
XX
XX Sequence 1726 BP; 284 A; 581 C; 499 G; 362 T; 0 U; 0 Other;
```

Alignment Scores:	9e-05	Length:	1726
Pred. No.:	16.00	Matches:	16
Score:	100.00%	Conservative:	0
Percent Similarity:	100.00%	Mismatches:	0
Best Local Similarity:	100.00%	Indels:	0
Query Match:	4.48%	Gaps:	0
DB:	6		

US-10-037-417-46 (1-357) X ABN85392 (1-1726)

Qy	1	MetAlaGlnLysGlyValLeuGlyProGlyGlnLeuGlyAlaValAla	16
Db	229	ATGGCCCAAGAGGGGGCTCTGGGGCTGGGCGAGCTGGGGGGCTGTGGCC	276

Search completed: February 25, 2004, 22:21:32
Job time : 431 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: February 25, 2004, 18:57:43 ; Search time 2496 Seconds
(without alignments)
4271.153 Million cell updates/sec

Title: US-10-037-417-46
Perfect score: 357
Sequence: 1 MAQGVGLPGQLGAVANS.....TKSLVLPWLSLHGLWGF 357

Scoring table: OLIGO
Xgapop 60.0 , Xgapext 60.0
Ygapop 60.0 , Ygapext 60.0
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 27513289 seqs, 14931090276 residues

Word size: 1
Total number of hits satisfying chosen parameters: 55025477

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Command line parameters: -MODEP=frame_p2n.model -DEV=xlp
-Q/cgn2_1/USPTO spool P/US10037417/runat 24022004 101108 18096/app query.fasta_1.519
-DB=EST -Qfmt=fastap -SUFFIX=olip2n.rst -MINMATCH=0.1 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=oligo -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR SCORE=quality -THR MIN=1 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc
-NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10037417@cgn 1 1 3549 @runat 24022004 101108 18096 -ICPU=3
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=60 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database :
EST:
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estnu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_estl:*
10: gb_estc2:*
11: gb_htc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: em_gss_hum:*
18: em_gss_inv:*
19: em_gss_pln:*
20: em_gss_vrt:*
21: em_gss_fun:*
22: em_gss_mam:*
23: em_gss_mus:*
24: em_gss_pro:*
25: em_gss_rod:*
26: em_gss_pbg:*
27: em_gss_vrl:*
28: gb_gssl:*

29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
C 1	154	43.1	537	10	AW450407 UI-H-BI3-
C 2	128	35.9	456	9	AI190509 qd49f07.x
C 3	57	16.0	930	13	EX436299 BX436299
C 4	35	9.8	296	10	BF561257 UI-R-CO-h
C 5	35	9.8	322	9	AA996811 UI-R-CO-h
C 6	23	6.4	351	13	BY236668 BY236668
C 7	23	6.4	365	13	BY235958 BY235958
C 8	23	6.4	369	13	BY236911 BY236911
C 9	23	6.4	430	28	AZ252377 RPCI-23-4
C 10	23	6.4	433	10	BB850607 BB850607
C 11	23	6.4	454	28	AZ875648 ZM0190113
C 12	20	5.6	433	10	BB850564 BB850564
C 13	17	4.8	875	13	BU164841 AGENCOURT
C 14	17	4.8	907	14	CD251990 AGENCOURT
C 15	17	4.8	929	13	BU860221 AGENCOURT
C 16	16	4.5	249	10	BB593614
C 17	16	4.5	294	9	AA300017
C 18	16	4.5	318	12	BM828821
C 19	16	4.5	360	9	AI393077
C 20	16	4.5	366	9	AI623099
C 21	16	4.5	377	9	AI193435
C 22	16	4.5	406	14	CB116948
C 23	16	4.5	409	9	AU298208
C 24	16	4.5	433	10	BB850795
C 25	16	4.5	462	9	AI343968
C 26	16	4.5	472	9	AI761519
C 27	16	4.5	479	9	AU298317
C 28	16	4.5	479	14	CB996389
C 29	16	4.5	492	10	BF063417
C 30	16	4.5	516	12	BM311864
C 31	16	4.5	527	9	AI658798
C 32	16	4.5	527	12	BG482000
C 33	16	4.5	534	9	AI393068
C 34	16	4.5	537	12	BM837284
C 35	16	4.5	537	12	BM837397
C 36	16	4.5	542	10	BE280845
C 37	16	4.5	555	12	BM837589
C 38	16	4.5	564	12	BM750655
C 39	16	4.5	570	12	BG479276
C 40	16	4.5	570	14	CB125988
C 41	16	4.5	575	9	AI207222
C 42	16	4.5	583	12	BG386531
C 43	16	4.5	590	13	BQ292219
C 44	16	4.5	591	9	AU134797
C 45	16	4.5	594	10	BE281529

ALIGNMENTS

RESULT 1
AW450407/c
LOCUS
DEFINITION
IMAGE:2735037 3', mRNA sequence.
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
1 (bases 1 to 537)
537 bp
mRNA
linear
EST 17-FEB-2000
UI-H-BI3-akn-g-11-0-UI.sl NCI CGAP Sub5 Homo sapiens cDNA clone
GI:6991183
Homo sapiens (human)
Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

AUTHORS NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
 TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 JOURNAL Tumor Gene Index
 COMMENT Unpublished (1997)
 Contact: Robert Strausberg, Ph.D.
 Email: cgapbs-r@mail.nih.gov
 Oligo-dr track not found, Not 1 site shown in beginning of sequence
 is likely internal to the message. cDNA Library Preparation: M.B.
 Soares Lab Clone distribution: NCI-CGAP clone distribution
 information can be found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbhp/image/image.html
 Seq primer: M13 Forward
 POLYA=No.

FEATURES source
 1..537
 Location/Qualifiers
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:2735037"
 /clone_lib="NCI CGAP Subs"
 /lab_host="DH10B (Life Technologies)"
 /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
 polylinker; Site 1: Not 1; Site 2: Eco RI; NCI CGAP Subs
 is a subtracted library derived from NCI CGAP Sub4. The
 NCI CGAP Subs library had 3 million recombinants. A
 single-stranded DNA preparation of NCI CGAP Sub4 was used
 as a tracer in a subtractive hybridization with a driver
 comprising: the IMAGE pool (NCI CGAP Kid3 pool 1 LLAM
 3334-3337, 3682-3683, 3798-3803) (IMAGE Clones)
 1322376-1323911, 1456008-1456775, 1500552-1502855);
 NCI CGAP Kid5 pool 1 LLAM 3338-3342, 3722-3725, 3776-3778
 (IMAGE Clones) 1323912-1325831, 1471368-1472903,
 1492104-1493255); NCI CGAP Lu5 pool 1 LLAM 3575-3582,
 3851-3854 (IMAGE Clones)
 1414920-1417991, 1520904-1522439); NCI CGAP GC4 pool 1 LLAM
 3164-3167, 3716-3720, 3733-3735 (IMAGE Clones)
 1257096-1258631, 1469064-1470983, 1475592-1476743);
 NCI CGAP Pr22 pool 1 LLAM 2457-2459, 2758-2759, 3062-3068
 (IMAGE Clones) 985608-986759, 1101192-1101959,
 1217928-1220615); NCI CGAP Co10 pool 1 LLAM 2644-2653,
 2871-2872 (IMAGE Clones)
 1057416-1061255, 1144584-1145351). (10% of the driver
 population), plus a pool of 3,840 arrayed clones from
 NCI CGAP Sub1 (IMAGE Clones) 2708616-2710535) and
 NCI CGAP Sub2 (IMAGE Clones) 2710536-2712455) (10% of
 the driver population), plus a pool of 11,136 clones from
 NCI CGAP Sub3 (IMAGE Clones) 2712456-2723591) (10% of the
 driver population), plus a pool of 5,472 clones from
 NCI CGAP Sub4 (IMAGE Clones) 2723592-2728969) (70% of the
 driver population). Subtraction was performed as
 previously described [Bonaldo, Lennon & Soares (1996):
 Normalization and Subtraction: Two Approaches To
 Facilitate Gene Discovery. Genome Research 6, 791-806.
 TAG TISSUE=colon
 TAG LIB=NCI CGAP_Co10
 TAG_SEQ=AAACG"

ORIGIN
 Alignment Scores:
 Pred. No.: 8,71e-127 Length: 537
 Score: 154.00 Matches: 154
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 43.14% Indels: 0
 DB: 10 Gaps: 0

US-10-037-417-46 (1-357) x AW450407 (1-537)

QY 100 GluTrpSerValLeuLeuGlyValHisSerGlnAspGlyProLeuAspGlyAlaHisThr 119
 Db 464 GAGTGGTGGTACTGCTGGCGGTGCACTCCAGGACGGCGCCCTGGACGGCGCACACC 405
 QY 120 ArgAlaValAlaAlaIleValProAlaAsnTyrSerGlnValGluLeuGlyAlaAsp 139

Db 404 CCGCAGTGGCGCCCATCGTGTGCGGCCCAACTACAGCCAGTGGAGTGGCGCCCGAC 345
 QY 140 LeuAlaLeuLeuArgLeuAlaSerProAlaSerLeuGlyProAlaValTrpProValCys 159
 Db 344 CTGGCCCTGCTGGCGCTGGCGCTCACCGCCAGCTGGGCCCCCGCGTGTGGCTGTCTGC 285
 QY 160 LeuProArgAlaSerHisArgPheValHisGlyThrAlaCysTrpAlaThrGlyTrpGly 179
 Db 284 CTGGCCCGCGCTCACCGCTTGTGACGCGCACCGCTGTGGGCCACCGGCTGGGGA 225
 QY 180 AspValGlnGluAlaAspProLeuProLeuProTrpValLeuGlnGluValGluLeuArg 199
 Db 224 GAGTCCAGGAGGAGATCTCTGCCTCTCCCTGGGTGTACAGGAAGTGGAGTAAGG 165
 QY 200 LeuLeuGlyGluAlaThrCysGlnCysLeuTyrSerGlnProGlyProPheAsnLeuThr 219
 Db 164 CTGTGGCGGAGGCCACCTGTCAATGTCTTACAGCCAGCCCGGTCCTTCACTCACT 105
 QY 220 LeuGlnIleLeuProGlyMetLeuCysAlaGlyTyrProGluGlyArgArgAspThrCys 239
 Db 104 CTCAGATATTGCCAGGAGTCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 45
 QY 240 GlnGlyAspSerGlyGlyProLeuValCysGlnGluGlyGly 253
 Db 44 CAGGCTGACTCTGGGGGGGCCCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 3

RESULT 2
 AL190509/c
 LOCUS
 DEFINITION Q449f07.x1 Soares fetal heart NDH19W Homo sapiens cDNA clone
 IMAGE:1732837 3' similar to SW:MCP6_MOUSE P21845 MAST CELL PROTEASE
 6 PRECURSOR ; mRNA sequence.
 AL190509
 VERSION AL190509.1 GI:3741718
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 456)
 NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
 National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 Tumor Gene Index
 Unpublished (1997)
 Contact: Robert Strausberg, Ph.D.
 Email: cgapbs-r@mail.nih.gov
 This clone is available royalty-free through LLNL ; contact the
 IMAGE Consortium (info@image.llnl.gov) for further information.
 Insert Length: 712 Std Error: 0.00
 Seq primer: -40UP from Gibco
 High quality sequence stop: 452.
 Location/Qualifiers
 1..456
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:1732837"
 /sex="unknown"
 /dev_stage="19 weeks"
 /lab_host="DH10B (ampicillin resistant)"
 /clone_lib="Soares fetal heart NDH19W"
 /note="Organ: heart; Vector: pT7T3D (Pharmacia) with a
 modified polylinker; Site 1: Not 1; Site 2: Eco RI; 1st
 strand cDNA was primed with a Not I - oligo(dT) primer [5',
 TGTACCAATCTGAAGTGGAGCGCGCATCTTTTTTTTTTTT 3'],
 double-stranded cDNA was size selected, ligated to Eco RI
 adapters (Pharmacia), digested with Not I and cloned into
 the Not I and Eco RI sites of a modified pT73 vector
 (Pharmacia). Library went through one round of
 normalization to a Cot = 5. Library constructed by
 M.Fatima Bonaldo. This library was constructed from the
 same fetus as the fetal lung library, Soares fetal lung

```

ORIGIN
NbHL19W."

Alignment Scores:
Pred. No.: 1.06e-103 Length: 456
Score: 128.00 Matches: 128
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 35.95% Indels: 0
DB: 9 Gaps: 0

US-10-037-417-46 (1-357) x A1190509 (1-456)
QY 121 AlaValAlaAlaValValProAlaAsnTyrSerGlnValGluLeuGlyAlaAspLeu 140
Db 384 GCAGTGGCCGCCATCTGGTGGCCGCCCACTACAGCAAGTGGAGCTGGGGCCGACCTG 325
QY 141 AlaLeuLeuArgLeuAlaSerProAlaSerLeuGlyProAlaValTyrProValCysLeu 160
Db 324 GCCTGTGCGCTGGCTCTACCGCCAGCTGGGCCCGCGTGTGGCTGTCTGCTG 265
QY 161 ProArgAlaSerHisArgPheValHisGlyThrAlaCysTyrAlaThrGlyTyrGlyAsp 180
Db 264 CCCGCGCCCTCAACCGCTTCGTGACGGCAGCCGCTGTGGCCACCGGCTGGGAGAC 205
QY 181 ValGlnGluAlaAspProLeuProLeuTyrProTyrValLeuGlnGluValGluLeuArgLeu 200
Db 204 GTCCAGGAGCAGATCTCTGCTCTCTCCCTGCTGCTACAGGAAGTGGAGCTTAAGGCTG 145
QY 201 LeuGlyGluAlaThrCysGlnCysLeuTyrSerGlnProGlyProPheAsnLeuThrLeu 220
Db 144 CTGGCGAGCCACCTGTCAATGCTCTACAGCAGCCGCTGCTTCACTCACTCTC 85
QY 221 GlnIleuProGlyMetLeuCysAlaGlyTyrProGluGlyArgAspThrCysGln 240
Db 84 CAGATATTCACAGGATGCTGTGTGCTGCTACCCAGAGCGCCGACGAGCAGCTGCCAG 25
QY 241 GlyAspSerGlyGlyProLeuVal 248
Db 24 GTGTACTCTGGGGGGCCCTGGTC 1

RESULT 3
BX436299
LOCUS
DEFINITION
3-PRIME mRNA sequence.
ACCESSION
BX436299
VERSION
BX436299.1 GI:30787522
KEYWORDS
EST.
SOURCE
Homo sapiens (human)
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
Full-length cDNA libraries and normalization
Unpublished (2001)
Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: segref@genoscope.cns.fr, Web : www.genoscope.cns.fr
Library was constructed by Life Technologies, a division of
Invitrogen. This sequence belongs to sequence cluster 7995.f For
more information about this cluster, see
http://www.genoscope.cns.fr/
cgi-bin/cluster.cgi?seq=CS0CAP001AD09FM1&cluster=7995.f. Contact :
Feng Liang Email : fliang@lifetech.com URL :
http://fulllength.invitrogen.com/ Invitrogen Corporation 1600
Faraday Avenue Genoscope sequence ID : CS0CAP001AD09FM1.
FEATURES
Location/Qualifiers
1..930
/organism="Homo sapiens"
/mol_type="mRNA"

ORIGIN
/db_xref="taxon:9606"
/clone="CS0CAP001YGI7"
/tissue_type="THYMUS"
/clone_lib="Homo sapiens THYMUS"
/note="Vector: pCMVSPORT 6; 1st strand cDNA was primed
with a NotI-oligo(dT) primer. Five prime end enriched,
double-strand cDNA was digested with NotI and cloned into
the NotI and EcoRV sites of the pCMVSPORT 6 vector.
Library was not normalized."

Alignment Scores:
Pred. No.: 4.21e-40 Length: 930
Score: 57.00 Matches: 57
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 15.97% Indels: 0
DB: 13 Gaps: 0

US-10-037-417-46 (1-357) x BX436299 (1-930)
QY 253 GlyArgTyrPheGlnAlaGlyIleThrSerPheGlyPheGlyCysGlyArgAsnArg 272
Db 70 GCGCGTGGTTCCAGGAGGAATCACAGCTTTGGCTTTGGCTGTGACGGAACCCG 129
QY 273 ProGlyValPheThrAlaValAlaThrTyrGluAlaTyrPheArgGluGlnValMetGly 292
Db 130 CTGGAGTTTTCACGTCTGTGCTACCTATGAGGCGATGATACGGAGCAGGTGTGGT 189
QY 293 SerGluProGlyProAlaPheProThrGlnProGlnLysThrGlnSerAsp 309
Db 190 TCAGAGCTGGGCTGCTTCCACCCAGCCAGACCCAGTGCAGAT 240

RESULT 4
BF561257/c
LOCUS
DEFINITION
UI-R-CO-hh-h-07-0-UI.r1 UI-R-CO Rattus norvegicus cDNA clone
ACCESSION
BF561257
VERSION
BF561257.1 GI:11670987
KEYWORDS
EST.
SOURCE
Rattus norvegicus (Norway rat)
ORGANISM
Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
1 (bases 1 to 296)
Ronald,M.F., Lennon,G. and Soares,M.B.
Normalization and subtraction: two approaches to facilitate gene
discovery
Genome Res. 6 (9), 791-806 (1996)
9704477
889548
Contact: Soares, MB
Coordinated Laboratory for Computational Genomics
University of Iowa
375 Newton Road, 4156 MEBRF, Iowa City, IA 52242, USA
Tel: 319 335 8250
Fax: 319 335 9565
Email: bento-soares@uiowa.edu
cDNA Library Preparation: M.B. Soares Lab Clone distribution:
clones will be available through Research Genetics (www.resgen.com)
This clone is also available through the I.M.A.G.E. Consortium at
LUNL (info@image.llnl.gov). IMAGE ID= 1781449
Seq primer: M13 Forward.
Location/Qualifiers
1..296
/organism="Rattus norvegicus"
/mol_type="mRNA"
/strain="Sprague-Dawley"
/db_xref="taxon:10116"
/clone="UI-R-CO-hh-h-07-0-UI"
/dev_stage="adult"

FEATURES
source

```

Fax: 319 335 9565
 Email: bento-soares@uiowa.edu
 Oligo-dt track not found, Not I site shown in beginning of sequence
 is likely internal to the message. cDNA Library Preparation: M.
 Fatima Bonaldo, Ph.D. Clone distribution: clones will be available
 through Research Genetics This clone is also available through the
 I.M.A.G.E. Consortium at LNL (info@image.llnl.gov). IMAGE
 ID=1781449
 Seq primer: M13 Forward
 POLYA=No.

Location/Qualifiers

1. 322
 /organism="Rattus norvegicus"
 /mol_type="mRNA"
 /strain="Sprague-Dawley"
 /db_xref="taxon:10116"
 /clones="UI-R-CO-hh-h-07-0-UI"
 /dev_stage="adult"
 /lab_host="DH10B (Life Technologies)"
 /clone_lib="UI-R-CO"
 /note="Vector: pT73D-Pac (Pharmacia) with a modified
 polylinker; Site 1: Not I; Site 2: Eco RI; The UI-R-CO
 library is a subtracted library derived from the UI-R-Al
 and UI-R-E1 libraries. The UI-R-Al library consisted of a
 mixture of individually tagged normalized libraries
 constructed from rat placenta, adult lung, brain, liver,
 kidney, heart, spleen, ovary, and muscle. The UI-R-E1
 library consisted of a mixture of individually tagged
 normalized libraries constructed from 8, 12 and 18-day
 embryo. The tag is a string of 3-5 nucleotides present
 between the Not I site and the oligo-dt track which
 allows identification of the library of origin of a clone
 within the mixture. The subtracted library (UI-R-CO) was
 constructed as follows: PCR amplified cDNA inserts from a
 pool of UI-R-Al and UI-R-E1 clones from which 3' ESTs had
 been derived was used as a driver in a hybridization with
 the pooled UI-R-Al and UI-R-E1 library in the form of
 single-stranded circles. The remaining single-stranded
 circles (subtracted library) was purified by
 hydroxyapatite column chromatography, converted to
 double-stranded circles and electroporated into DH10B
 bacteria (Life Technologies) to generate the UI-R-CO
 library. This procedure has been previously described
 (Bonaldo, Lennon and Soares, Genome Research 6: 791-806,
 1996)"

FEATURES

source

ORIGIN

Alignment Scores: 4.65e-21 Length: 296
 Pred. No.: 35.00 Matches: 35
 Score: 35.00
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 9.80% Indels: 0
 DB: 10 Gaps: 0

US-10-037-417-46 (1-357) x BF561257 (1-296)

QY 55 ProGlyThrTrpProTrrpGlnValSerLeuHisGlyGlyHisCysGlyGly 74
 Db 270 CCTGGCACTTGGCCGTGGAGTGCATCATCAGGTGAGGCCACATCTGCGGGGC 211
 QY 75 SerLeuLeuAlaProSerTrpValSerAlaHisCysPhe 89
 Db 210 TCCCTCATCGCCCTTCTCGGTCTCTCCGCTGCTACTGTTTC 166

RESULT 5
 AA996811
 LOCUS
 DEFINITION
 UI-R-CO-hh-h-07-0-UI.s1 UI-R-CO Rattus norvegicus cDNA clone
 UI-R-CO-hh-h-07-0-UI 3', mRNA sequence.

AA996811.1 GI:4280575

EST.

Rattus norvegicus (Norway rat)

ORGANISM

Normalisation and subtraction: two approaches to facilitate gene discovery

Genome Res. 6 (9), 791-806 (1996)

97044477

8889548

On Jun 5, 1998 this sequence version replaced gi:3187366.

Contact: Soares, MB

Coordinated Laboratory for Computational Genomics

University of Iowa

375 Newton Road, 4156 MEBRF, Iowa City, IA 52242, USA

Tel: 319 335 8250

ORIGIN

Alignment Scores: 5.11e-21 Length: 322
 Pred. No.: 35.00 Matches: 35
 Score: 35.00
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 9.80% Indels: 0
 DB: 9 Gaps: 0

US-10-037-417-46 (1-357) x AA996811 (1-322)

QY 55 ProGlyThrTrpProTrrpGlnValSerLeuHisGlyGlyHisCysGlyGly 74
 Db 50 CCTGGCACTTGGCCGTGGAGTGCATCATCAGGTGAGGCCACATCTGCGGGGC 109

QY 75 SerLeuLeuAlaProSerTrpValSerAlaHisCysPhe 89

Db 110 TCCCTCATCGCCCTTCTCGGTCTCTCCGCTGCTACTGTTTC 154

RESULT 6

BY236668

LOCUS

DEFINITION

CDNA clone F930049D04 5', mRNA sequence.

ACCESSION

BY236668

VERSION

BY236668.1

GI:26417846

KEYWORDS

BY236668 351 bp mRNA linear EST 10-DEC-2002
 BY236668 RIKEN full-length enriched, adult inner ear Mus musculus
 CDNA clone F930049D04 5', mRNA sequence.

SOURCE
Mus musculus (house mouse)
ORGANISM
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus;
1 (bases 1 to 351)
REFERENCE
AUTHORS
Okazaki, Y., Furuno, M., Kasukawa, T., Adachi, J., Bono, H., Kondo, S.,
Nikaido, I., Osato, N., Saito, R., Suzuki, H., Yamanaka, I.,
Kiyosawa, H., Yagi, K., Tomaru, Y., Hasegawa, Y., Nogami, A.,
Schonbach, C., Gojobori, T., Baldarelli, R., Hill, D.P., Bult, C.,
Batalov, S., Beisel, K.W., Blake, J.A., Bradt, D., Bruscia, H.,
Chothia, C., Corbani, L.E., Cousins, S., Dalla, E., Dragani, T.A.,
Fletcher, C.F., Forrest, A., Frazer, K.S., Gaasterland, T.,
Gariboldi, M., Gissi, C., Godzik, A., Gough, J., Grimmond, S.,
Gustincich, S., Hirokawa, N., Jackson, I.J., Jarvis, E.D., Kanai, A.,
Kawaji, H., Kawasawa, Y., Kedzierski, R.M., King, B.L., Konagaya, A.,
Kurochkin, I.V., Lee, Y., Lenhard, B., Lyons, P.A., Maglott, D.R.,
Maltais, L., Marchionni, L., McKenzie, L., Miki, H., Nagashima, T.,
Numata, K., Okido, T., Pavan, W.J., Pertea, G., Pesole, G.,
Petrovsky, N., Pillai, R., Pontius, J.U., Qi, D., Ramachandran, S.,
Ravasi, T., Reed, J.C., Reid, J., Ring, B.Z., Ringwald, M.,
Sandelin, A., Schneider, C., Sempie, C.A., Setou, M., Shimada, K.,
Sultana, R., Takenaka, Y., Taylor, M.S., Teasdale, R.D., Tomita, M.,
Verardo, R., Wagner, L., Wahlestedt, C., Wang, Y., Watanabe, Y.,
Wells, C., Wilming, L.G., Wynshaw-Boris, A., Yang, Y., Yanagisawa, M.,
Yang, L., Yuan, Z., Zavolan, M., Zhu, Y., Zimmer, A., Carninci, P.,
Hayatsu, N., Hirozane-Kishikawa, T., Konno, H., Nakamura, M.,
Sakazume, N., Sato, K., Shiraki, T., Waki, K., Kawai, J., Aizawa, K.,
Arakawa, T., Fukuda, S., Hara, A., Hashizume, W., Imotani, K., Ishii, Y.,
Itoh, M., Kagawa, I., Miyazaki, A., Sakai, K., Sasaki, D., Shibata, K.,
Shinagawa, A., Yasunishi, A., Yoshino, M., Waterston, R., Lander, E.S.,
Rogers, J., Birney, E. and Hayashizaki, Y.
TITLE
Analysis of the mouse transcriptome based on functional annotation
of 60,770 full-length cDNAs
JOURNAL
Nature 420, 563-573 (2002)
MEDLINE
22354683
PUBMED
12456651
COMMENT
Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic
Sciences Center (GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@gsc.riken.go.jp
URL: <http://genome.gsc.riken.go.jp/>
Aizawa, K., Akimura, T., Arakawa, T., Carninci, P., Fukuda, S.,
Hirozane, T., Imotani, K., Ishii, Y., Itoh, M., Kawai, J., Konno, H.,
Miyazaki, A., Murata, M., Nakamura, M., Nomura, K., Numazaki, R.,
Ohno, M., Sakai, K., Sakazume, N., Sasaki, D., Sato, K., Shibata, K.,
Shiraki, T., Tagami, M., Waki, K., Watabiki, A., Muramatsu, M. and
Hayashizaki, Y. Direct Submission
Computational Analysis of Full-length Mouse cDNAs Compared with
Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)
Normalization and subtraction of cap-trapper-selected cDNAs to
prepare full-length cDNA libraries for rapid discovery of new
genes. Genome Res. 10 (10), 1617-1630 (2000)
RIKEN integrated sequence analysis (RISA) system-384-format
sequencing pipeline with 384 multicapillary sequencer. Genome Res.
10 (11), 1757-1771 (2000)
Computer-based methods for the mouse full-length cDNA
encyclopedia: real-time sequence clustering for construction of a
nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
cDNA library was prepared and sequenced in Mouse Genome
Encyclopedia Project of Genome Exploration Research Group in RIKEN
Genomic Sciences Center and Genome Science Laboratory in RIKEN.
Division of Experimental Animal Research in Riken contributed to
prepare mouse tissues.
Tissues were provided by Kirk W. Beisel (Boys Town National
Research Hospital 555 North 30th Street Omaha, NE 68131 USA) whose
assistance we gratefully acknowledge. Please visit our web site
(<http://genome.gsc.riken.go.jp/>) for further details.
Location/Qualifiers

source
1. .351
/organism="Mus musculus"
/mol_type="mRNA"
/strain="CS7BL/6J"
/db_xref="taxon:10090"
/clone="F930049D04"
/tissue_type="inner ear"
/dev_stage="adult"
/clone_lib="RIKEN full-length enriched, adult inner ear"
ORIGIN
Alignment Scores:
Pred. No.: 2,76e-10 Length: 351
Score: 23.00 Matches: 23
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 6.44% Indels: 0
DB: 13 Gaps: 0
US-10-037-417-46 (1-357) x BY236668 (1-351)
QY 67 GlyGlyGlyHisIleCysGlyGlySerLeuIleAlaProSerTrpValLeuSerAlaAla 86
Db 268 GTGGGGGGCCACATCTGCGGGGGCTCCCTCATCGCCCTTCTCGTCTCGTCTCGTCT 327
QY 87 HisCysPhe 89
Db 328 CACTCTTTC 336
RESULT 7
BY235958
LOCUS
BY235958 RIKEN full-length enriched, adult inner ear Mus musculus
cDNA clone F930045I01 5', mRNA sequence.
365 bp mRNA
EST.
BY235958.1 GI:26417082
Mus musculus (house mouse)
ORGANISM
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus;
1 (bases 1 to 365)
REFERENCE
AUTHORS
Okazaki, Y., Furuno, M., Kasukawa, T., Adachi, J., Bono, H., Kondo, S.,
Nikaido, I., Osato, N., Saito, R., Suzuki, H., Yamanaka, I.,
Kiyosawa, H., Yagi, K., Tomaru, Y., Hasegawa, Y., Nogami, A.,
Schonbach, C., Gojobori, T., Baldarelli, R., Hill, D.P., Bult, C.,
Batalov, S., Beisel, K.W., Blake, J.A., Bradt, D., Bruscia, H.,
Chothia, C., Corbani, L.E., Cousins, S., Dalla, E., Dragani, T.A.,
Fletcher, C.F., Forrest, A., Frazer, K.S., Gaasterland, T.,
Gariboldi, M., Gissi, C., Godzik, A., Gough, J., Grimmond, S.,
Gustincich, S., Hirokawa, N., Jackson, I.J., Jarvis, E.D., Kanai, A.,
Kawaji, H., Kawasawa, Y., Kedzierski, R.M., King, B.L., Konagaya, A.,
Kurochkin, I.V., Lee, Y., Lenhard, B., Lyons, P.A., Maglott, D.R.,
Maltais, L., Marchionni, L., McKenzie, L., Miki, H., Nagashima, T.,
Numata, K., Okido, T., Pavan, W.J., Pertea, G., Pesole, G.,
Petrovsky, N., Pillai, R., Pontius, J.U., Qi, D., Ramachandran, S.,
Ravasi, T., Reed, J.C., Reid, J., Ring, B.Z., Ringwald, M.,
Sandelin, A., Schneider, C., Sempie, C.A., Setou, M., Shimada, K.,
Sultana, R., Takenaka, Y., Taylor, M.S., Teasdale, R.D., Tomita, M.,
Verardo, R., Wagner, L., Wahlestedt, C., Wang, Y., Watanabe, Y.,
Wells, C., Wilming, L.G., Wynshaw-Boris, A., Yang, Y., Yanagisawa, M.,
Yang, L., Yuan, Z., Zavolan, M., Zhu, Y., Zimmer, A., Carninci, P.,
Hayatsu, N., Hirozane-Kishikawa, T., Konno, H., Nakamura, M.,
Sakazume, N., Sato, K., Shiraki, T., Waki, K., Kawai, J., Aizawa, K.,
Arakawa, T., Fukuda, S., Hara, A., Hashizume, W., Imotani, K., Ishii, Y.,
Itoh, M., Kagawa, I., Miyazaki, A., Sakai, K., Sasaki, D., Shibata, K.,
Shinagawa, A., Yasunishi, A., Yoshino, M., Waterston, R., Lander, E.S.,
Rogers, J., Birney, E. and Hayashizaki, Y.
TITLE
Analysis of the mouse transcriptome based on functional annotation
of 60,770 full-length cDNAs
JOURNAL
Nature 420, 563-573 (2002)
MEDLINE
22354683

12466851
 PUBLISHED
 COMMENT
 Contact: Yoshihide Hayashizaki
 Laboratory for Genome Exploration Research Group, RIKEN Genomic
 Sciences Center(GSC), Yokohama Institute
 The Institute of Physical and Chemical Research (RIKEN)
 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
 Tel: 81-45-503-9222
 Fax: 81-45-503-9216
 Email: genome-res@gsc.riken.go.jp/
 URL: http://genome.gsc.riken.go.jp/
 Aizawa, K., Akimura, T., Arakawa, T., Carninci, P., Fukuda, S.,
 Hirozane, T., Imotani, K., Ishii, Y., Itoh, M., Kawai, J., Konno, H.,
 Miyazaki, A., Murata, M., Nakamura, M., Nomura, K., Numazaki, R.,
 Ohno, M., Sakai, K., Sakazume, N., Sasaki, D., Sato, K., Shibata, K.,
 Shiraki, T., Tagami, M., Waki, K., Watahiki, A., Muramatsu, M. and
 Hayashizaki, Y. Direct Submission
 Computational Analysis of Full-length Mouse cDNAs Compared with
 Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)
 Normalization and subtraction of cap-trapper-selected cDNAs to
 prepare full-length cDNA libraries for rapid discovery of new
 genes. Genome Res. 10 (10), 1617-1630 (2000)
 RIKEN integrated sequence analysis (RISA) system--384-format
 sequencing pipeline with 384 multicapillary sequencer. Genome Res.
 10 (11), 1757-1771 (2000)
 Computer-based methods for the mouse full-length cDNA
 encyclopedia: real-time sequence clustering for construction of a
 nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
 cDNA library was prepared and sequenced in Mouse Genome
 Encyclopedia Project of Genome Exploration Research Group in Riken
 Genomic Sciences Center and Genome Science Laboratory in RIKEN.
 Division of Experimental Animal Research in Riken contributed to
 prepare mouse tissues.
 Tissues were provided by Kirk W. Beisel (Boys Town National
 Research Hospital 555 North 30th Street Omaha, NE 68131 USA) whose
 assistance we gratefully acknowledge. Please visit our web site
 (http://genome.gsc.riken.go.jp) for further details.
 Location/Qualifiers
 source 1. 365
 /organism="Mus musculus"
 /mol_type="mRNA"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone="F930045I01"
 /tissue_type="inner ear"
 /dev_stage="adult"
 /clone_lib="RIKEN full-length enriched, adult inner ear"
 ORIGIN
 Alignment Scores:
 Pred. No.: 2.88e-10 Length: 365
 Score: 23.00 Matches: 23
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 6.44% Indels: 0
 DB: 13 Gaps: 0
 US-10-037-417-46 (1-357) x BY235958 (1-365)
 QY 67 GlyGlyGlyHisIleCysGlyGlySerIleuIleAlaProSerTrpValIleuSerIleAla 86
 Db 265 GTGGGGGCCCATCTGCGGGGGCTCCCTCATCGCCCTTCTCGGTCTCTCCGCTGCT 324
 QY 87 HisCysPhe 89
 Db 325 CACTGTTC 333
 RESULT 8
 BY236911
 LOCUS
 DEFINITION
 BY236911 RIKEN full-length enriched, adult inner ear Mus musculus
 CDNA clone F930050D15 5', mRNA sequence.
 ACCESSION
 BY236911
 VERSION
 BY236911.1 GI:26418099

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

EST.

Mus musculus (house mouse)

Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 369)
 Okazaki, Y., Furuno, M., Kasukawa, T., Adachi, J., Bono, H., Kondo, S.,
 Nikaido, I., Osato, N., Saito, R., Suzuki, H., Yamanaka, T.,
 Kiyosawa, H., Yagi, K., Tomaru, Y., Hasegawa, Y., Nogami, A.,
 Schonbach, C., Gojobori, T., Baldarelli, R., Hill, D. P., Bult, C.,
 Hume, D. A., Quackenbush, J., Schriml, L. M., Kanapin, A., Matsuda, H.,
 Batalov, S., Beisel, K. W., Blake, J. A., Bradt, D., Brusic, V.,
 Chothia, C., Corbani, L. E., Cousins, S., Dalia, E., Dragani, T. A.,
 Fletcher, C. F., Forrest, A., Frazer, K. S., Gaasterland, T.,
 Gariboldi, M., Glasi, C., Godzik, A., Gough, J., Grimmond, S.,
 Gustincich, S., Hirokawa, N., Jackson, I. J., Jarvis, E. D., Kanai, A.,
 Kawai, H., Kawasawa, Y., Kedzierzki, R. M., King, B. L., Konagaya, A.,
 Kurochkin, I. V., Lee, Y., Lenhard, B., Lyons, P. A., Maglott, D. R.,
 Maltais, L., Marchionni, L., McKenzie, L., Miki, H., Nagashima, T.,
 Numata, K., Okido, T., Pavan, W. J., Pertea, G., Resole, G.,
 Petrovsky, N., Pillai, R., Pontius, J. U., Qi, D., Ramachandran, S.,
 Ravasi, T., Reed, J. C., Reed, D. J., Reid, J., Ring, B. Z., Ringwald, M.,
 Sandelin, A., Schneider, C., Semple, C. A., Setou, M., Shimada, K.,
 Sultana, R., Takenaka, Y., Taylor, M. S., Teasdale, R. D., Tomita, M.,
 Verardo, R., Wagner, L., Wahlestedt, C., Wang, Y., Watanabe, Y.,
 Wells, C., Wilming, L. G., Wynshaw-Boris, A., Yanagisawa, M., Yang, I.,
 Yang, L., Yuan, Z., Zavolan, M., Zhu, Y., Zimmer, A., Carninci, P.,
 Hayatsu, N., Hirozane-Kishikawa, T., Konno, H., Nakamura, M.,
 Sakazume, N., Sato, K., Shiraki, T., Waki, K., Kawai, J., Aizawa, K.,
 Arakawa, T., Fukuda, S., Hara, A., Hashizume, W., Imotani, K., Ishii, Y.,
 Itoh, M., Kagawa, I., Miyazaki, A., Sakai, K., Sasaki, D., Shibata, K.,
 Shinagawa, A., Yasunishi, A., Yoshino, M., Waterston, R., Lander, E. S.,
 Rogers, J., Birney, E. and Hayashizaki, Y.
 Analysis of the mouse transcriptome based on functional annotation
 of 60,770 full-length cDNAs
 Nature 420, 563-573 (2002)
 12466851
 COMMENT
 Contact: Yoshihide Hayashizaki
 Laboratory for Genome Exploration Research Group, RIKEN Genomic
 Sciences Center(GSC), Yokohama Institute
 The Institute of Physical and Chemical Research (RIKEN)
 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
 Tel: 81-45-503-9222
 Fax: 81-45-503-9216
 Email: genome-res@gsc.riken.go.jp/
 URL: http://genome.gsc.riken.go.jp/
 Aizawa, K., Akimura, T., Arakawa, T., Carninci, P., Fukuda, S.,
 Hirozane, T., Imotani, K., Ishii, Y., Itoh, M., Kawai, J., Konno, H.,
 Miyazaki, A., Murata, M., Nakamura, M., Nomura, K., Numazaki, R.,
 Ohno, M., Sakai, K., Sakazume, N., Sasaki, D., Sato, K., Shibata, K.,
 Shiraki, T., Tagami, M., Waki, K., Watahiki, A., Muramatsu, M. and
 Hayashizaki, Y. Direct Submission
 Computational Analysis of Full-length Mouse cDNAs Compared with
 Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)
 Normalization and subtraction of cap-trapper-selected cDNAs to
 prepare full-length cDNA libraries for rapid discovery of new
 genes. Genome Res. 10 (10), 1617-1630 (2000)
 RIKEN integrated sequence analysis (RISA) system--384-format
 sequencing pipeline with 384 multicapillary sequencer. Genome Res.
 10 (11), 1757-1771 (2000)
 Computer-based methods for the mouse full-length cDNA
 encyclopedia: real-time sequence clustering for construction of a
 nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
 cDNA library was prepared and sequenced in Mouse Genome
 Encyclopedia Project of Genome Exploration Research Group in Riken
 Genomic Sciences Center and Genome Science Laboratory in RIKEN.
 Division of Experimental Animal Research in Riken contributed to
 prepare mouse tissues.
 Tissues were provided by Kirk W. Beisel (Boys Town National
 Research Hospital 555 North 30th Street Omaha, NE 68131 USA) whose
 assistance we gratefully acknowledge. Please visit our web site
 (http://genome.gsc.riken.go.jp) for further details.

Thu Feb 26 13:50:26 2004

us-10-037-417-46.olip2n.rst

FEATURES
source
Location/Qualifiers
1. .369
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="F930050D15"
/tissue_type="inner ear"
/dev_stage="adult"
/clone_lib="RIKEN full-length enriched, adult inner ear"

ORIGIN

Alignment Scores:
Pred. No.: 2,92e-10 Length: 369
Score: 23.00 Matches: 23
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 6.44% Indels: 0
DB: 13 Gaps: 0

US-10-037-417-46 (1-357) x BY236911 (1-369)

Qy 67 GlyGlyHisIleCysGlySerLeuIleAlaProSerTrpValLeuSerAlaAla 86
Db 271 GGTGGGGCCACATCTGGGGGGCTCCCTCATCGCCCTTCTGGGTCTCTCCGCTGCT 330
Qy 87 HisCysPhe 89
Db 331 CACTGTTTC 339

RESULT 9

AZ252377
LOCUS AZ252377 430 bp DNA linear GSS 15-JUN-2000
DEFINITION RPCI-23-466N15.TVB RPCI-23 Mus musculus genomic clone
RPCI-23-466N15, genomic survey sequence.

ACCESSION

AZ252377

GSS.

Mus musculus (house mouse)

Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 430)

Zhao, S., Nierman, W., Feidlyum, T., Malek, J., Shatsman, S.,

Akiret, B., Levins, M., McGann, S., Tsegaye, G., Geer, K., Krol, M., de

Jong, P. and Fraser, C.M.

Mouse BAC End Sequences from Library RPCI-23

Unpublished (1999)

Other GSSs: RPCI-23-466N15.TJ RPCI-23-466N15.TV RPCI-23-466N15.TVB

Contact: Shaying Zhao

Department of Eukaryotic Genomics

The Institute for Genomic Research

9712 Medical Center Dr., Rockville, MD 20850, USA

Tel: 301 838 0200

Fax: 301 838 0208

Email: szhao@tigr.org

Clones are derived from the mouse BAC library RPCI-23. For BAC

library availability, please contact Pieter de Jong

(pieter@dejong.med.buffalo.edu). Clones may be purchased from

BACPAC Resources (http://bacpac.med.buffalo.edu/orderingframe.html)

or from Resea ch Genetics (info@resgen.com). BAC end page:

http://www.tigr.org/tdb/bac_ends/mouse/bac_end_intro.html

Plate: 466 row: N column: 15

Seq primer: SP6

Class: BAC ends.

Location/Qualifiers

1. .430

/organism="Mus musculus"

/mol_type="genomic DNA"

/strain="C57BL/6J"

/db_xref="taxon:10090"

/clone="RPCI-23-466N15"

/sex="Female"

FEATURES

source

/lab_host="DH10B"
/clone_lib="RPCI-23"
/note="Organ: Kidney/Brain; Vector: pBACE3.6; Site 1:
EcoRI; Site 2: EcoRI; Female C57BL/6J mouse kidney and/or
brain genomic DNA was isolated and partially digested
with a combination of EcoRI and EcoRI Methyase. Size
selected DNA was cloned into the pBACE3.6 vector at the
EcoRI sites. The ligation products were transformed into
DH10B electrocompetent cells (BRL Life Technologies)."

ORIGIN

Alignment Scores:
Pred. No.: 3,46e-10 Length: 430
Score: 23.00 Matches: 23
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 6.44% Indels: 0
DB: 28 Gaps: 0

US-10-037-417-46 (1-357) x AZ252377 (1-430)

Qy 258 AlaGlyIleThrSerPheGlyPheGlyCysGlyArgGlnArgProGlyValpHeThr 277
Db 93 GCCGGAATCACCAGCTTTGGCTTTGGCTTGTGGACGAGGACCGCCCTGGGGTCTTCACT 152
Qy 278 AlaValAla 280
Db 153 GCAGTGGCT 161

RESULT 10

BB850607

LOCUS

DEFINITION

BB850607 RIKEN full-length enriched, adult inner ear Mus musculus

CDNA clone F930108E12 5', mRNA sequence.

ACCESSION BB850607

VERSION BB850607.1 GI:17092061

KEYWORDS EST.

SOURCE Mus musculus (house mouse)

ORGANISM

REFERENCE

AUTHORS

1 (bases 1 to 433)

Akimura, T., Arakawa, T., Carninci, P., Furuno, M., Hanagaki, T.,

Hayatsu, N., Hiramoto, K., Hiraoka, T., Hirozane, T., Imotani, K.,

Ishii, Y., Ito, M., Kawai, J., Kojima, Y., Konno, H., Kouda, M.,

Matsuyama, T., Nakamura, M., Nishi, K., Nomura, K., Numasaki, R.,

Okazaki, Y., Okido, T., Saito, R., Sakai, K., Sakai, K., Sakazume, N.,

Sasaki, D., Sato, K., Shibata, K., Shinagawa, A., Shiraki, T.,

Sogabe, Y., Suzuki, H., Tagawa, A., Takahashi, F., Takaku-Akahira, S.,

Tanaka, T., Tonaru, A., Toya, T., Watahiki, A., Yasunishi, A.,

Muramatsu, M. and Hayashizaki, Y.

RIKEN Encyclopedia of Mouse Full-length cDNAs (Akimura, T., et al.

2001)

TITLE

JOURNAL

COMMENT

Unpublished (2001)

Contact: Yoshihide Hayashizaki

Laboratory for Genome Exploration Research Group, RIKEN Genomic

Sciences Center (GSC), Yokohama Institute

The Institute of Physical and Chemical Research (RIKEN)

1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan

Tel: 81-45-503-9222

Fax: 81-45-503-9216

Email: genome-res@sc.riken.go.jp,

URL: http://genome.gsc.riken.go.jp/

Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,

Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.

Normalization and subtraction of cap-trapper-selected cDNAs to

prepare full-length cDNA libraries for rapid discovery of new

genes. Genome Res. 10 (10), 1617-1630 (2000)

wagi, K., Fujiwaka, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E.,

Watahiki, M., Itoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T.,

Matsuura, S., Kawai, Y., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A.

and Hayashizaki, Y.

RIKEN integrated sequence analysis (RISA) system--384-format

sequencing pipeline with 384 multicapillary sequencer. Genome Res. 10 (11) 1757-1771 (2000)
 Kono,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P., Sugahara,Y. and Hayashizaki,Y.
 Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
 Please visit our web site (<http://genome.gsc.riken.go.jp>) for further details.

e mouse tissues.

FEATURES

Location/Qualifiers
 1. .433
 /organism="Mus musculus"
 /mol_type="mRNA"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone="F930108E12"
 /tissue_type="inner ear"
 /dev_stage="adult"
 /clone_lib="RIKEN full-length enriched, adult inner ear"

ORIGIN

Alignment Scores:
 Pred. No.: 3.49e-10 Length: 433
 Score: 23.00 Matches: 23
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 6.44% Indels: 0
 DB: 10 Gaps: 0

US-10-037-417-46 (1-357) x BB850607 (1-433)

QY 67 GlyGlyGlyHisIleCysGlyGlySerLeuIleAlaProSerTrpValLeuSerAlaAla 86
 Db 274 GGTGGGGGGGCACATCTCGGGGGGCTCCCTCATGGCCCTCTCGGTCTCTCCGTGCT 333
 QY 87 HisCysPhe 89
 Db 334 CACTGTTC 342

RESULT 11
 AZ875648/c 454 bp DNA linear GSS 21-FEB-2001
 LOCUS
 DEFINITION
 2M0190113F Mouse 10kb plasmid UUGC1M library Mus musculus genomic clone UUGC2M0190113 F, genomic survey sequence.

ACCESSION
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM

Mus musculus (house mouse)
 Mus musculus
 Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 454)
 Dunn,B., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C., Iellam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly,M., Rose,R., Stokes,R., Tangey,A., von Niederhausern,A. and Wright,D.,Weiss,R.
 Mouse whole genome scaffolding with paired end reads from 10kb plasmid inserts

Unpublished (2000)
 Contact: Robert B. Weiss
 University of Utah Genome Center
 University of Utah
 Rm. 309, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLG, UT 84112, USA
 Tel: 801 585 5606
 Fax: 801 585 7177
 Email: ddunn@genetics.utah.edu
 Insert Length: 10000 Std Error: 0.00
 Plate: 0190 row: 1 column: 13
 Seq primer: CGTTGTAACGACGGCCAGT
 Claes: plasmid ends
 High quality sequence stop: 454.

FEATURES

Location/Qualifiers
 1. .454
 /organism="Mus musculus"
 /mol_type="genomic DNA"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone="UUGC2M0190113"
 /sex="Male"
 /lab_host="E. Coli strain XL10-Gold, Tl-resistant, F-"
 /clone_lib="Mouse 10kb plasmid UUGC1M library"
 /note="Vector: PWD42nv; Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource
 (<http://www.jax.org/resources/documents/dnares/>). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptor DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of PWD42 (gi|4732114|gb|AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptor mouse DNA was annealed to adaptor vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

ORIGIN

Alignment Scores:
 Pred. No.: 3.68e-10 Length: 454
 Score: 23.00 Matches: 23
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 6.44% Indels: 0
 DB: 28 Gaps: 0

US-10-037-417-46 (1-357) x AZ875648 (1-454)

QY 67 GlyGlyGlyHisIleCysGlyGlySerLeuIleAlaProSerTrpValLeuSerAlaAla 86
 Db 271 GGTGGGGGGGCACATCTCGGGGGGCTCCCTCATGGCCCTCTCGGTCTCTCCGTGCT 212
 QY 87 HisCysPhe 89
 Db 211 CACTGTTC 203

RESULT 12

BB850564
 LOCUS
 DEFINITION
 BB850564 RIKEN full-length enriched, adult inner ear Mus musculus cDNA clone F930108A18 5', mRNA sequence.

ACCESSION
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM

Mus musculus (house mouse)
 Mus musculus
 Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 433)

REFERENCE
 AUTHORS
 Akimura,T., Arakawa,T., Carninci,P., Furuno,M., Hanagaki,T., Hayatsu,N., Hiramoto,K., Hirooka,T., Hirozane,T., Imotani,K., Ishii,Y., Ito,M., Kawai,J., Kojima,Y., Konno,H., Kouda,M., Matsuyama,T., Nakamura,M., Nishi,K., Nomura,K., Numasaki,R., Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K., Sakazume,N., Sasaki,D., Sato,K., Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H., Tagawa,A., Takahashi,F., Takaku-Akaira,S., Tanaka,T., Tomaru,A., Toya,T., Watanishi,A., Yasunishi,A., Muramatsu,M. and Hayashizaki,Y.
 RIKEN Encyclopedia of Mouse Full-length cDNAs (Akimura,T., et al. 2001)

JOURNAL
COMMENT

Unpublished (2001)
Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic
Sciences Center (GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-Ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@gsc.riken.go.jp,
URL: http://genome.gsc.riken.go.jp/
Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,
Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M., and Hayashizaki, Y.
Normalization and subtraction of cap-trapper-selected cDNAs to
prepare full-length cDNA libraries for rapid discovery of new
genes. *Genome Res.* 10 (10): 1617-1630 (2000)
wagi, K., Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E.,
Watahiki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T.,
Matsura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A.,
and Hayashizaki, Y.
RIKEN integrated sequence analysis (RISA) system--384-format
sequencing pipeline with 384 multicapillary sequencer. *Genome Res.*
10 (11): 1757-1771 (2000)
Konno, H., Fukunishi, Y., Shibata, K., Itoh, M., Carninci, P.,
Sugahara, Y., and Hayashizaki, Y.
Computer-based methods for the mouse full-length cDNA
encyclopedia: real-time sequence clustering for construction of a
nonredundant cDNA library. *Genome Res.* 11 (2): 281-289 (2001)
Please visit our web site (<http://genome.gsc.riken.go.jp>) for
further details.
e mouse tissues.

FEATURES
source

1. .433
Location/Qualifiers
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="F930108A18"
/tissue_type="inner ear"
/dev_stage="adult"
/clone_lib="RIKEN full-length enriched, adult inner ear"

ORIGIN

Alignment Scores:
Pred. No.: 1.64e-07 Length: 433
Score: 20.00 Matches: 20
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 5.60% Indels: 0
DB: 10 Gaps: 0

US-10-037-417-46 (1-357) x BB850564 (1-433)

QY 67 GYGlyGlyHisIleCysGlyGlySerLeulleAlaProSerrTpValIeuSerAlaAla 86
Db 274 GTGGGGCCACATCTGCGGGGCTCCCTCATCGCCCTCTCGGCTCTCTCGCTCTCT 333

RESULT 13
BUI64841/c

LOCUS BUI64841 875 bp mRNA linear EST 04-SEP-2002
DEFINITION AGENCOURT_7906304 NTH_MGC_67 Homo sapiens cDNA clone IMAGE:6140519
5', mRNA sequence.
ACCESSION BUI64841
VERSION BUI64841.1 GI:22678793
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens

REFERENCE
1 (bases 1 to 875)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
AUTHORS NIH-MGC <http://mgs.nci.nih.gov/>
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.

Email: cgabbs-remail.nih.gov

Tissue Procurement: ATCC
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
<http://image.llnl.gov>
Plate: LHAM3459 row: d column: 24
High quality sequence stop: 645.

FEATURES
source

1. .875
Location/Qualifiers
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:6140519"
/tissue_type="retinoblastoma"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH MGC 67"
/notes="Organ: eye; Vector: pCMV-SPORT6; Site: 1: NotI;
Site: 2: SalI; Cloned unidirectionally. Primer: Oligo dr.
Average insert size 1.75 kb. Library constructed by Life
Technologies."

ORIGIN

Alignment Scores:
Pred. No.: 0.00017 Length: 875
Score: 17.00 Matches: 17
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 4.76% Indels: 0
DB: 13 Gaps: 0

US-10-037-417-46 (1-357) x BUI64841 (1-875)

QY 341 LeuValLeuProTrrpLeuSerProHisSerLeuLeuGlyLeuTrrpGlyPhe 357
Db 668 CTTGTCTCTCCCTGGCTCTCTCCACACTCTCTCTGGGCTCTGGGGTTT 618

RESULT 14
CD251990/c

LOCUS CD251990 907 bp mRNA linear EST 22-MAY-2003
DEFINITION AGENCOURT_14211954 NIH_MGC_180 Homo sapiens cDNA clone
IMAGE:30383555 5', mRNA sequence.
ACCESSION CD251990
VERSION CD251990.1 GI:31012456
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 907)
NIH-MGC <http://mgs.nci.nih.gov/>.
AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Daniela S. Gerhard, Ph.D.
Office of Cancer Genomics
National Cancer Institute / NIH
Bldg. 31 Rm10A07 Bethesda, MD 20892
Email: cgabbs-remail.nih.gov
Tissue Procurement: Dr. Michael Brownstein
cDNA Library Preparation: Invitrogen Corp
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
<http://image.llnl.gov>
Plate: NDAM447 row: j column: 12
High quality sequence stop: 548.

FEATURES
source

1. .907
Location/Qualifiers
/organism="Homo sapiens"
/mol_type="mRNA"

/db_xref=taxon:9606"
/clone=IMAGE:30383555"
/lab_host="DH10B-tcn A (T1 and T5 phage resistances)"
/clone_lib="NIH_MGC_180"
/note="Organ: Testis; Vector: pCMV-SPORT6.1; Site_1: NotI;
Site_2: EcoRV (destroyed); Library is oligo-dT primed and
directionally cloned (EcoRV site is destroyed upon
cloning). Average insert size 1.68 kb. Library was
constructed by (Invitrogen). Note: this is a NIH_MGC
Library."

ORIGIN

Alignment Scores:

Pred. No.:	0.000177	Length:	907
Score:	17.00	Matches:	17
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	4.76%	Indels:	0
DB:	14	Gaps:	0

US-10-037-417-46 (1-357) x CD2511990 (1-907)

QY 341 LeuValLeuProTrpLeuSerProHisSerLeuGlyLeuTrpGlyPhe 357
Db 520 CTTGTCTCCCGGCTCTCTCCACACTCTCTCTGGGCTCTGGGGGTTT 470

RESULT 15

BUB860221/c
LOCUS
DEFINITION AGENCOURT 10435504 NIH_MGC_107 Homo sapiens cDNA clone
IMAGE:6651279 5', mRNA sequence.
ACCESSION BUB860221
VERSION BUB860221.1 GI:24045213
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 929)
AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs@mail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone Distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
Plate: L1CM2897 row: n column: 15
High quality sequence stop: 423.
Location/Qualifiers
1. .929
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone=IMAGE:6651279"
/tissue_type="adenocarcinoma, cell line"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH_MGC_107"
/note="Organ: breast; Vector: pOTB7; Site 1: EcoRI;
Site 2: XhoI; cDNA made by oligo-dT priming.
Directionally cloned into EcoRV/XhoI sites using the
following 5' adaptor: GGCACGAG(G). Library constructed by
Ling Hong in the laboratory of Gerald M. Rubin (University
of California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies).
Note: this is a NIH_MGC Library."

FEATURES

source

1. .929
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone=IMAGE:6651279"
/tissue_type="adenocarcinoma, cell line"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH_MGC_107"
/note="Organ: breast; Vector: pOTB7; Site 1: EcoRI;
Site 2: XhoI; cDNA made by oligo-dT priming.
Directionally cloned into EcoRV/XhoI sites using the
following 5' adaptor: GGCACGAG(G). Library constructed by
Ling Hong in the laboratory of Gerald M. Rubin (University
of California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies).
Note: this is a NIH_MGC Library."

ORIGIN

Alignment Scores:

Pred. No.:	0.000182	Length:	929
Score:	17.00	Matches:	17
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	4.76%	Indels:	0
DB:	13	Gaps:	0

US-10-037-417-46 (1-357) x BUB860221 (1-929)

QY 341 LeuValLeuProTrpLeuSerProHisSerLeuGlyLeuTrpGlyPhe 357
Db 520 CTTGTCTCCCGGCTCTCTCCACACTCTCTCTGGGCTCTGGGGGTTT 2

Search completed: February 26, 2004, 00:06:29
Job time : 2502 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - nucleic search, using frame_plus.p2n model

Run on: February 25, 2004, 18:59:12 ; Search time 97 Seconds

(without alignments)

2042.448 Million cell updates/sec

Title: US-10-037-417-46

Perfect score: 357

Sequence: 1 MAQKGVLPGLGAVANSDS.....TKSLVLPWLSPLHLLGWF 357

Scoring table:

OLIGO
Xgapop 60.0 , Xgapext 60.0
Ygapop 60.0 , Ygapext 60.0
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 682709 seqs, 277475446 residues

Word size: 1

Total number of hits satisfying chosen parameters: 1360453

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Command line parameters:

-MODEL=frame.p2n.model -DEV=xlp
-Q=/cgn2_1/USPTO.spool.p/US10037417/runat 24022004 101109 18111/app query.fasta_1.519
-DB=Issued Patents NA -QFMT=fastap -SUFFIX=olip2n.rni -MINWATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=oligo -TRANS=human40.cdi
-LIST=45 -DOCALLIGN=200 -THR SCORE=quality -THR MIN=1 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10037417 @CGN 1 1 103 @runat 24022004 101109 18111 -NCPU=6 -ICPU=3
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSRBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database : Issued Patents NA:*

1: /cgn2_6/ptodata/2/ina/5A COMB.seq:*\n2: /cgn2_6/ptodata/2/ina/5B COMB.seq:*\n3: /cgn2_6/ptodata/2/ina/6A COMB.seq:*\n4: /cgn2_6/ptodata/2/ina/6B COMB.seq:*\n5: /cgn2_6/ptodata/2/ina/PTCUS COMB.seq:*\n6: /cgn2_6/ptodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	13	3.6	506	4	US-09-280-116-72
2	13	3.6	792	4	US-09-244-111-3
3	13	3.6	944	3	US-09-070-526-1
4	13	3.6	994	3	US-09-008-271A-19
5	13	3.6	1049	4	US-09-386-642-9
6	13	3.6	1343	4	US-09-618-259-72
7	13	3.6	1360	4	US-09-618-259-6
8	11	3.1	262	3	US-08-944-483-5
9	11	3.1	340	4	US-09-280-116-28
10	11	3.1	340	4	US-09-280-116-29
11	11	3.1	436	3	US-08-906-769-122
12	11	3.1	436	3	US-08-906-616-122

13	11	3.1	436	3	US-08-639-075A-122	Sequence 122, App
14	11	3.1	436	3	US-09-012-431-122	Sequence 122, App
15	11	3.1	436	3	US-09-032-215-19	Sequence 19, Appl
16	11	3.1	436	3	US-09-012-692-122	Sequence 122, App
17	11	3.1	436	3	US-08-906-613-122	Sequence 122, App
18	11	3.1	453	6	5200340-3	Patent No. 5200340
19	11	3.1	662	4	US-09-702-705-109	Sequence 109, App
20	11	3.1	662	4	US-09-736-457-109	Sequence 109, App
21	11	3.1	662	4	US-09-614-124B-109	Sequence 109, App
22	11	3.1	662	4	US-09-671-325-109	Sequence 109, App
23	11	3.1	662	4	US-09-589-184-109	Sequence 109, App
24	11	3.1	690	1	US-08-379-621-1	Sequence 1, Appli
25	11	3.1	690	1	US-08-889-078-1	Sequence 1, Appli
26	11	3.1	732	2	US-08-361-395-2	Sequence 2, Appli
27	11	3.1	735	3	US-09-079-970A-1	Sequence 1, Appli
28	11	3.1	771	3	US-09-079-970A-4	Sequence 4, Appli
29	11	3.1	833	2	US-08-790-137-2	Sequence 2, Appli
30	11	3.1	840	4	US-09-244-111-5	Sequence 5, Appli
31	11	3.1	933	4	US-09-023-942A-29	Sequence 29, Appli
32	11	3.1	1031	2	US-08-978-404B-1	Sequence 1, Appli
33	11	3.1	1052	4	US-09-386-642-10	Sequence 10, Appli
34	11	3.1	1065	1	US-08-427-640-1	Sequence 1, Appli
35	11	3.1	1065	2	US-08-427-640-5	Sequence 5, Appli
36	11	3.1	1065	2	US-08-811-949-60	Sequence 60, Appli
37	11	3.1	1068	1	US-08-137-116-2	Sequence 2, Appli
38	11	3.1	1068	1	US-08-427-640-3	Sequence 3, Appli
39	11	3.1	1068	1	US-08-427-640-7	Sequence 7, Appli
40	11	3.1	1068	2	US-08-811-949-44	Sequence 44, Appli
41	11	3.1	1068	2	US-08-811-949-46	Sequence 46, Appli
42	11	3.1	1068	2	US-08-811-949-52	Sequence 52, Appli
43	11	3.1	1068	2	US-08-811-949-58	Sequence 58, Appli
44	11	3.1	1068	6	523256-3	Patent No. 523256
45	11	3.1	1081	2	US-09-016-366A-22	Sequence 22, Appli

ALIGNMENTS

RESULT 1

US-09-280-116-72/c

; Sequence 72, Application US/09280116A

; Patent No. 6331427

; GENERAL INFORMATION:

; APPLICANT: Robison, Keith E.

; TITLE OF INVENTION: Nucleic Acid Molecules Encoding Human Protease Homologs

; FILE REFERENCE: 5800-24, 035800/176965

; CURRENT APPLICATION NUMBER: US/09/280,116A

; CURRENT FILING DATE: 1999-03-26

; NUMBER OF SEQ ID NOS: 268

; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 72

; LENGTH: 506

; TYPE: DNA

; ORGANISM: Homo sapiens

; FEATURE:

; OTHER INFORMATION: trypsin-like serine proteases

; FEATURE:

; NAME/KEY: misc feature

; LOCATION: (1)-(506)

; OTHER INFORMATION: n = a, t, c or g

US-09-280-116-72

Alignment Scores:
Pred. No.: 0.000809
Score: 13.00
Length: 506
Matches: 13
Percent Similarity: 100.00%
Best Local Similarity: 100.00%
Query Match: 3.64%
DB: 4
Conservative: 0
Mismatch: 0
Indels: 0
Gaps: 0

US-10-037-417-46 (1-357) x US-09-280-116-72 (1-506)

Qy 239 CysGlnGlyAspSerGlyProLeuValCysGlu 251
|||||

Db 329 TCCAGGGTGATTCGGAGGCCGCTGGTGTGGAGGA 291
RESULT 2
US-09-244-111-3
; Sequence 3, Application US/09244111
; Patent No. 6566498
; GENERAL INFORMATION:
; APPLICANT: Ni, et al.
; TITLE OF INVENTION: Human Serine Protease and Serpin Polypeptides
; FILE REFERENCE: PF391
; CURRENT APPLICATION NUMBER: US/09/244,111
; EARLIER FILING DATE: 1999-02-04
; EARLIER APPLICATION NUMBER: 60/073,961
; EARLIER FILING DATE: 1998-02-06
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 3
; LENGTH: 792
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (85)..(708)
US-09-244-111-3
Alignment Scores:
Pred. No.: 0.00125 Length: 792
Score: 13.00 Matches: 13
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 3.64% Indels: 0
DB: 4 Gaps: 0
US-10-037-417-46 (1-357) x US-09-244-111-3 (1-792)
QY 237 AspThrCysGlnGlyAspSerGlyGlyProLeuValCys 249
Db 541 GACACGTGCCAGGCGATTCTGGAGGCCCTCGTGTGT 579
RESULT 3
US-09-070-526-1
; Sequence 1, Application US/09070526
; Patent No. 6100059
; GENERAL INFORMATION:
; APPLICANT: SOUTHAN, CHRISTOPHER
; APPLICANT: CLINKENBEARD, HELEN
; APPLICANT: BURGESS, NICOLA
; TITLE OF INVENTION: NO. 6100059e1 Compounds
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: RATNER & PRESTIA
; STREET: P.O. BOX 980
; CITY: VALLEY FORGE
; STATE: PA
; COUNTRY: USA
; ZIP: 19482
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FASTSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/070,526
; FILING DATE: 30-APR-1998
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 9711952.3
; FILING DATE: 9-JUN-1997
; APPLICATION NUMBER: EP 97309646.4
; FILING DATE: 1-DEC-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: PRESTIA, PAUL F
; REGISTRATION NUMBER: 23,031

; REFERENCE/DOCKET NUMBER: GH-30353
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 610-407-0700
; TELEFAX: 610-407-0701
; TELEX: 846169
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 944 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
US-09-070-526-1
Alignment Scores:
Pred. No.: 0.00149 Length: 944
Score: 13.00 Matches: 13
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 3.64% Indels: 0
DB: 3 Gaps: 0
US-10-037-417-46 (1-357) x US-09-070-526-1 (1-944)
QY 237 AspThrCysGlnGlyAspSerGlyGlyProLeuValCys 249
Db 720 GACACGTGCCAGGCGATTCTGGAGGCCCTCGTGTGT 758
RESULT 4
US-09-008-271A-19
; Sequence 19, Application US/09008271A
; Patent No. 6203979
; GENERAL INFORMATION:
; APPLICANT: Bandman, Olga
; Hillman, Jennifer L.
; Yue, Henry
; Guegler, Karl J.
; Corley, Neil C.
; Tang, Tom Y.
; Shah, Purvi
; TITLE OF INVENTION: HUMAN PROTEASE MOLECULES
; NUMBER OF SEQUENCES: 24
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Dr.
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/008,271A
; FILING DATE: 16-Jan-1998
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: <Unknown>
; FILING DATE: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Mohan-Peterson, Sheela
; REGISTRATION NUMBER: 41,201
; REFERENCE/DOCKET NUMBER: PF-0458 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650-855-0555
; TELEFAX: 650-845-4166
; INFORMATION FOR SEQ ID NO: 19:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 994 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear


```
; IMMEDIATE SOURCE:
; LIBRARY: COLNROT27
; CLONE: 1798496
; SEQUENCE DESCRIPTION: SEQ ID NO: 19 :
US-09-008-271A-19

Alignment Scores:
Pred. No.: 0.00156 Length: 994
Score: 13.00 Matches: 13
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 3.64% Indels: 0
DB: 3 Gaps: 0

US-10-037-417-46 (1-357) x US-09-008-271A-19 (1-994)

Qy 237 AspThrCysGlnGlyAspSerGlyClyProLeuValCys 249
Db 770 GACACGTGCCAGGCGGATCTGGAGGCCCGCCCTGGTGTGT 808

RESULT 5
US-09-386-642-9
; Sequence 9, Application US/09386642
; Patent No. 6420157
; GENERAL INFORMATION:
; APPLICANT: Darrow, Andrew
; APPLICANT: Qi, Jensen
; APPLICANT: Andrade-Gordon, Patricia
; TITLE OF INVENTION: Zymogen Activation System
; FILE REFERENCE: ORT-1028
; CURRENT APPLICATION NUMBER: US/09/386,642
; CURRENT FILING DATE: 1999-08-31
; NUMBER OF SEQ ID NOS: 60
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 9
; LENGTH: 1049
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Fusion gene
; OTHER INFORMATION: with homo sapien serine protease catalytic domain
US-09-386-642-9

Alignment Scores:
Pred. No.: 0.00165 Length: 1049
Score: 13.00 Matches: 13
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 3.64% Indels: 0
DB: 4 Gaps: 0

US-10-037-417-46 (1-357) x US-09-386-642-9 (1-1049)

Qy 237 AspThrCysGlnGlyAspSerGlyClyProLeuValCys 249
Db 688 GACACGTGCCAGGCGGATCTGGAGGCCCGCCCTGGTGTGT 726

RESULT 6
US-09-618-259-72/c
; Sequence 72, Application US/09618259
; Patent No. 6642013
; GENERAL INFORMATION:
; APPLICANT: O'Brien, Timothy J.
; APPLICANT: Underwood, Lowell J.
; TITLE OF INVENTION: No. 6642013el Extracellular Serine Protease
; FILE REFERENCE: D6020CIP2
; CURRENT APPLICATION NUMBER: US/09/618,259
; CURRENT FILING DATE: 2000-07-18
; PRIOR APPLICATION NUMBER: US 09/127,444
; PRIOR FILING DATE: 1998-08-21
; NUMBER OF SEQ ID NOS: 72
; SEQ ID NO 72
; LENGTH: 1343
```

```
; TYPE: DNA
; ORGANISM: artificial sequence
; FEATURE:
; OTHER INFORMATION: Anti-sense sequence of TADG-14
US-09-618-259-72

Alignment Scores:
Pred. No.: 0.0021 Length: 1343
Score: 13.00 Matches: 13
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 3.64% Indels: 0
DB: 4 Gaps: 0

US-10-037-417-46 (1-357) x US-09-618-259-72 (1-1343)

Qy 237 AspThrCysGlnGlyAspSerGlyClyProLeuValCys 249
Db 213 GACACGTGCCAGGCGGATCTGGAGGCCCGCCCTGGTGTGT 175

RESULT 7
US-09-618-259-6
; Sequence 6, Application US/09618259
; Patent No. 6642013
; GENERAL INFORMATION:
; APPLICANT: O'Brien, Timothy J.
; APPLICANT: Underwood, Lowell J.
; TITLE OF INVENTION: No. 6642013el Extracellular Serine Protease
; FILE REFERENCE: D6020CIP2
; CURRENT APPLICATION NUMBER: US/09/618,259
; CURRENT FILING DATE: 2000-07-18
; PRIOR APPLICATION NUMBER: US 09/127,444
; PRIOR FILING DATE: 1998-08-21
; NUMBER OF SEQ ID NOS: 72
; SEQ ID NO 6
; LENGTH: 1360
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: Nucleotide sequence encoding Tumor Antigen Derived
; OTHER INFORMATION: Gene-14 (TADG-14) protein; nt 1344-1360 NCEI accession
; OTHER INFORMATION: #AA343629
US-09-618-259-6

Alignment Scores:
Pred. No.: 0.00213 Length: 1360
Score: 13.00 Matches: 13
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 3.64% Indels: 0
DB: 4 Gaps: 0

US-10-037-417-46 (1-357) x US-09-618-259-6 (1-1360)

Qy 237 AspThrCysGlnGlyAspSerGlyClyProLeuValCys 249
Db 1131 GACACGTGCCAGGCGGATCTGGAGGCCCGCCCTGGTGTGT 1169

RESULT 8
US-08-944-483-5
; Sequence 5, Application US/08944483
; Patent No. 6232456
; GENERAL INFORMATION:
; APPLICANT: COHEN, MAURICE
; APPLICANT: COLPITTS, TRACEY L.
; APPLICANT: FRIEDMAN, PAULA N.
; APPLICANT: GRANADOS, EDWARD N.
; APPLICANT: KLASS, MICHAEL R.
; APPLICANT: RUSSELL, JOHN C.
; APPLICANT: STEWART, KENT D.
; APPLICANT: STROUPE, STEVEN D.
; TITLE OF INVENTION: NOVEL SERINE PROTEASE REAGENTS
; TITLE OF INVENTION: AND METHODS USEFUL FOR DETECTING AND TREATING DISEASES
```

```

; TITLE OF INVENTION: OF THE PROSTATE
; NUMBER OF SEQUENCES: 76
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Abbott Laboratories
; STREET: 100 Abbott Park Road
; CITY: Abbott Park
; STATE: IL
; COUNTRY: USA
; ZIP: 60064-3500
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/944,483
; FILING DATE:
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Becker, Cheryl L.
; REGISTRATION NUMBER: 35,441
; REFERENCE/DOCKET NUMBER: 6183.US.01
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 847/935-1729
; TELEFAX: 847/938-2623
; TELEX:
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 262 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-944-483-5

Alignment Scores:
Pred. No.: 0.0407 Length: 262
Score: 11.00 Matches: 11
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 3.08% Indels: 0
DB: 3 Gaps: 0

US-10-037-417-46 (1-357) x US-08-944-483-5 (1-262)
QY 239 CysGlnGlyAspSerGlyProLeuValCys 249
DB 70 TGCCAGGGTGACTCCGGGGGCCCTCTGCTGT 102

RESULT 9
US-09-280-116-28/c
; Sequence 28, Application US/09280116A
; Patent No. 6331427
; GENERAL INFORMATION:
; APPLICANT: Robison, Keith E.
; TITLE OF INVENTION: Nucleic Acid Molecules Encoding Human Protease Homologs
; FILE REFERENCE: 5800-24, 035800/176965
; CURRENT APPLICATION NUMBER: US/09/280,116A
; CURRENT FILING DATE: 1999-03-26
; NUMBER OF SEQ ID NOS: 268
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 28
; LENGTH: 340
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: trypsin-like serine proteases
US-09-280-116-28

Alignment Scores:
Pred. No.: 0.0525 Length: 340
Score: 11.00 Matches: 11
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 3.08% Indels: 0
DB: 4 Gaps: 0

US-10-037-417-46 (1-357) x US-09-280-116-28 (1-340)
QY 240 GlnGlyAspSerGlyProLeuValCysGlu 250
DB 250 CAGGAGATTCTGGGGGCCCTCTGTGTGAA 218

RESULT 10
US-09-280-116-29/c
; Sequence 29, Application US/09280116A
; Patent No. 6331427
; GENERAL INFORMATION:
; APPLICANT: Robison, Keith E.
; TITLE OF INVENTION: Nucleic Acid Molecules Encoding Human Protease Homologs
; FILE REFERENCE: 5800-24, 035800/176965
; CURRENT APPLICATION NUMBER: US/09/280,116A
; CURRENT FILING DATE: 1999-03-26
; NUMBER OF SEQ ID NOS: 268
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 29
; LENGTH: 340
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: trypsin-like serine proteases
US-09-280-116-29

Alignment Scores:
Pred. No.: 0.0525 Length: 340
Score: 11.00 Matches: 11
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 3.08% Indels: 0
DB: 4 Gaps: 0

US-10-037-417-46 (1-357) x US-09-280-116-29 (1-340)
QY 240 GlnGlyAspSerGlyProLeuValCysGlu 250
DB 250 CAGGAGATTCTGGGGGCCCTCTGTGTGAA 218

RESULT 11
US-08-906-769-122
; Sequence 122, Application US/08906769
; Patent No. 6077687
; GENERAL INFORMATION:
; APPLICANT: Grieve, Robert B.
; APPLICANT: Rushlow, Keith E.
; APPLICANT: Wu Hunter, Shirley
; APPLICANT: Frank, Glenn R.
; APPLICANT: Stiegler, Gary
; APPLICANT: Gaines, Patrick J.
; APPLICANT: Silver, Gary
; TITLE OF INVENTION: FLUA PROTEASE PROTEINS, NUCLEIC ACID
; TITLE OF INVENTION: MOLECULES AND USES THEREOF
; NUMBER OF SEQUENCES: 190
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sheridan Ross & McIntosh
; STREET: 1700 Lincoln Street, Suite 3500
; CITY: Denver
; STATE: Colorado
; COUNTRY: USA
; ZIP: 80203
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
```

```
/
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: US/08/906,769
/ FILING DATE:
/
/ CLASSIFICATION:
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER: US 08/639,075
/ FILING DATE: 24-APR-1996
/
/ ATTORNEY/AGENT INFORMATION:
/ NAME: Connell, Gary J.
/ REGISTRATION NUMBER: 32,020
/ REFERENCE/DOCKET NUMBER: 2618-25-C2
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: (303) 863-9700
/ TELEFAX: (303) 863-0223
/ INFORMATION FOR SEQ ID NO: 122:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 436 base pairs
/ TYPE: nucleic acid
/ STRANDEDNESS: single
/ TOPOLOGY: linear
/ MOLECULE TYPE: cdna
/ FEATURE:
/ NAME/KEY: CDS
/ LOCATION: 2..436
/ OTHER INFORMATION: /note= "At pos. bp 301, change A to
/ OTHER INFORMATION: W; at pos. bp 342, change C to Y; at pos. bp 397, change C to
/ OTHER INFORMATION: at pos. bp 431, change G to S. At pos. aa 100, 114 and 144,
/ OTHER INFORMATION: substitute Xaa."
/
/ US-08-906-769-122
/
/ Alignment Scores:
/ Pred. No.: 0.0669 Length: 436
/ Score: 11.00 Matches: 11
/ Percent Similarity: 100.00% Conservative: 0
/ Best Local Similarity: 100.00% Mismatches: 0
/ Query Match: 3.08% Indels: 0
/ DB: 3 Gaps: 0
/
/ US-10-037-417-46 (1-357) x US-08-906-769-122 (1-436)
/
/ Qy 237 AspThrCysGlnGlyAspSerGlyGlyProLeu 247
/ Db 257 GACACTTCCAGGAGATTTCAGGTGGCCATTG 289
/
/ RESULT 12
/ US-08-906-616-122
/ Sequence 122, Application US/08906616
/ Patent No. 6121035
/ GENERAL INFORMATION:
/ APPLICANT: Grieve, Robert B.
/ APPLICANT: Rushlow, Keith E.
/ APPLICANT: Wu Hunter, Shirley
/ APPLICANT: Frank, Glenn R.
/ APPLICANT: Stiegler, Gary
/ APPLICANT: Gaines, Patrick J.
/ TITLE OF INVENTION: FLEA AMINOPEPTIDASE PROTEINS AND USES THEREOF
/ NUMBER OF SEQUENCES: 190
/ CORRESPONDENCE ADDRESSES:
/ ADDRESSEE: Sheridan Ross P.C.
/ STREET: 1700 Lincoln Street, Suite 3500
/ CITY: Denver
/ STATE: Colorado
/ COUNTRY: USA
/ ZIP: 80203
/ COMPUTER READABLE FORM:
/ MEDIUM TYPE: Floppy disk
/ OPERATING SYSTEM: IBM PC compatible
/ SOFTWARE: PatentIn Release #1.0, Version #1.30
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: US/08/906,616
/ FILING DATE: 05-AUG-1997
/
/
/
/
/ CLASSIFICATION: 536
/ ATTORNEY/AGENT INFORMATION:
/ NAME: Connell, Gary J.
/ REGISTRATION NUMBER: 32,020
/ REFERENCE/DOCKET NUMBER: 2618-25-C2-3
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: (303) 863-9700
/ TELEFAX: (303) 863-0223
/ INFORMATION FOR SEQ ID NO: 122:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 436 base pairs
/ TYPE: nucleic acid
/ STRANDEDNESS: single
/ TOPOLOGY: linear
/ MOLECULE TYPE: cdna
/ FEATURE:
/ NAME/KEY: CDS
/ LOCATION: 2..436
/ OTHER INFORMATION: /note= "At pos. bp 301, change A to
/ OTHER INFORMATION: W; at pos. bp 342, change C to Y; at pos. bp 397, change C to
/ OTHER INFORMATION: at pos. bp 431, change G to S. At pos. aa 100, 114 and 144,
/ OTHER INFORMATION: substitute Xaa."
/
/ US-08-906-616-122
/
/ Alignment Scores:
/ Pred. No.: 0.0669 Length: 436
/ Score: 11.00 Matches: 11
/ Percent Similarity: 100.00% Conservative: 0
/ Best Local Similarity: 100.00% Mismatches: 0
/ Query Match: 3.08% Indels: 0
/ DB: 3 Gaps: 0
/
/ US-10-037-417-46 (1-357) x US-08-906-616-122 (1-436)
/
/ Qy 237 AspThrCysGlnGlyAspSerGlyGlyProLeu 247
/ Db 257 GACACTTCCAGGAGATTTCAGGTGGCCATTG 289
/
/ RESULT 13
/ US-08-639-075A-122
/ Sequence 122, Application US/08639075A
/ Patent No. 6150125
/ GENERAL INFORMATION:
/ APPLICANT: Grieve, Robert B.
/ APPLICANT: Rushlow, Keith E.
/ APPLICANT: Wu Hunter, Shirley
/ APPLICANT: Frank, Glenn R.
/ APPLICANT: Stiegler, Gary
/ APPLICANT: Gaines, Patrick J.
/ TITLE OF INVENTION: FLEA PROTEASE PROTEINS, NUCLEIC ACID
/ NUMBER OF SEQUENCES: 190
/ CORRESPONDENCE ADDRESSES:
/ ADDRESSEE: Sheridan Ross & McIntosh
/ STREET: 1700 Lincoln Street, Suite 3500
/ CITY: Denver
/ STATE: Colorado
/ COUNTRY: USA
/ ZIP: 80203
/ COMPUTER READABLE FORM:
/ MEDIUM TYPE: Floppy disk
/ OPERATING SYSTEM: IBM PC compatible
/ SOFTWARE: PatentIn Release #1.0, Version #1.30
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: US/08/639,075A
/ FILING DATE: 24-APR-1996
/ CLASSIFICATION: 424
/ ATTORNEY/AGENT INFORMATION:
/ NAME: Connell, Gary J.
/ REGISTRATION NUMBER: 32,020
/ REFERENCE/DOCKET NUMBER: 2618-25-C2
```

```
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: (303) 863-9700
/ TELEFAX: (303) 863-0223
/ INFORMATION FOR SEQ ID NO: 122:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 436 base pairs
/ TYPE: nucleic acid
/ STRANDEDNESS: single
/ TOPOLOGY: linear
/ MOLECULE TYPE: cdna
/ FEATURE:
/ NAME/KEY: CDS
/ LOCATION: 2..436
/ OTHER INFORMATION: /note= "at pos. bp 301, change A to
/ W; at pos. bp 342, change C to Y; at pos. bp 397, change C to Y;
/ at pos. bp 431, change G to S. At pos. aa 100, 114 and 144,
/ substitute Xaa."
/ SEQUENCE DESCRIPTION: SEQ ID NO: 122:
US-08-639-075A-122
US-10-037-417-46 (1-357) x US-08-639-075A-122 (1-436)

Alignment Scores:
Pred. No.: 0.0669 Length: 436
Score: 11.00 Matches: 11
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 3.08% Indels: 0
DB: 3 Gaps: 0

US-10-037-417-46 (1-357) x US-09-012-431-122 (1-436)
QY 237 AspThrCysGlnGlyAspSerGlyGlyProLeu 247
DB 257 GACACTTCCAGGAGATTTCAGGTGGGCCATTG 289

RESULT 14
US-09-012-431-122
; Sequence 122, Application US/09012431
; Patent No. 6180383
; GENERAL INFORMATION:
; APPLICANT: Grieve, Robert B.
; Rushlow, Keith E.
; Wu Hunter, Shirley
; Frank, Glenn R.
; Stiegler, Gary
; Gaines, Patrick J.
; Silver, Gary
; TITLE OF INVENTION: FLEA PROTEASE PROTEINS, NUCLEIC ACID
; MOLECULES AND USES THEREOF
; NUMBER OF SEQUENCES: 190
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sheridan Ross & McIntosh
; STREET: 1700 Lincoln Street, Suite 3500
; CITY: Denver
; STATE: Colorado
; COUNTRY: USA
; ZIP: 80203
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/012,431
; FILING DATE: 23-Jan-1998
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/639,075
; FILING DATE: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Connell, Gary J.
; REGISTRATION NUMBER: 32,020
; REFERENCE/DOCKET NUMBER: 2618-25-C2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (303) 863-9700

/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: (303) 863-9700
/ TELEFAX: (303) 863-0223
/ INFORMATION FOR SEQ ID NO: 122:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 436 base pairs
/ TYPE: nucleic acid
/ STRANDEDNESS: single
/ TOPOLOGY: linear
/ MOLECULE TYPE: cdna
/ FEATURE:
/ NAME/KEY: CDS
/ LOCATION: 2..436
/ OTHER INFORMATION: /note= "at pos. bp 301, change A to
/ W; at pos. bp 342, change C to Y; at pos. bp 397, change C to Y;
/ at pos. bp 431, change G to S. At pos. aa 100, 114 and 144,
/ substitute Xaa."
/ SEQUENCE DESCRIPTION: SEQ ID NO: 122:
US-09-012-431-122
US-10-037-417-46 (1-357) x US-09-012-431-122 (1-436)

Alignment Scores:
Pred. No.: 0.0669 Length: 436
Score: 11.00 Matches: 11
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 3.08% Indels: 0
DB: 3 Gaps: 0

US-10-037-417-46 (1-357) x US-09-012-431-122 (1-436)
QY 237 AspThrCysGlnGlyAspSerGlyGlyProLeu 247
DB 257 GACACTTCCAGGAGATTTCAGGTGGGCCATTG 289

RESULT 15
US-09-032-215-19
; Sequence 19, Application US/09032215
; Patent No. 6204010
; GENERAL INFORMATION:
; APPLICANT: Stiegler, Gary L.
; APPLICANT: Gaines, Patrick J.
; TITLE OF INVENTION: FLEA PROTEASE PROTEINS, NUCLEIC
; TITLE OF INVENTION: ACID MOLECULES, AND USES THEREOF
; NUMBER OF SEQUENCES: .50
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sheridan Ross P.C.
; STREET: 1700 Lincoln Street, Suite 3500
; CITY: Denver
; STATE: Colorado
; COUNTRY: U.S.A.
; ZIP: 80203
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: ASCII DOS TEXT
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/032,215
; FILING DATE: 27-FEB-1998
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Connell, Gary J.
; REGISTRATION NUMBER: 32,020
; REFERENCE/DOCKET NUMBER: 2618-25-C6
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (303) 863-9700
; TELEFAX: (303) 863-0223
; INFORMATION FOR SEQ ID NO: 19:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 436 nucleotides
; TYPE: nucleic acid
; STRANDEDNESS: single
```

```

; TOPOLOGY: linear
; MOLECULE TYPE: CDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..436
; FEATURE:
; NAME/KEY: W = A
; LOCATION: 301
; FEATURE:
; NAME/KEY: Y = C
; LOCATION: 342, 397
; FEATURE:
; NAME/KEY: S = C
; LOCATION: 431
; FEATURE:
; NAME/KEY: Xaa = Unknown
; LOCATION: 100, 114, 144
; US-09-032-215-19

```

```

Alignment Scores:
Pred. No.: 0.0669 Length: 436
Score: 11.00 Matches: 11
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 3.08% Indels: 0
DB: 3 Gaps: 0

```

US-10-037-417-46 (1-357) x US-09-032-215-19 (1-436)

```

Qy 237 AspThrCysGlnGlyAspSerGlyGlyProLeu 247
Db 257 GACACTGCCAGGGAGATTCAGGTGGGCATTG 289

```

Search completed: February 26, 2004, 00:08:19
Job time : 101 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: February 25, 2004, 23:28:21 ; Search time 399 Seconds
(without alignments)

3133.050 Million cell updates/sec

Title: US-10-037-417-46

Perfect score: 357

Sequence: 1 MAQGVLPQQLGAVANSDS.....TKSLVLPWLSPHSLGLWGF 357

Scoring table:

OLIGO	
Xgapop 60.0 , Xgapext 60.0	
Ygapop 60.0 , Ygapext 60.0	
Fgapop 6.0 , Fgapext 7.0	
Delop 6.0 , Delext 7.0	

Searched: 2308684 seqs, 1750822206 residues

Word size: 1

Total number of hits satisfying chosen parameters: 4610896

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Command line parameters:

-WORDLEN=frame+ p2n.model -DEV=xlp
-Q=/cgn2_1/USPTO_spool/p/US10037417/runat_24022004_101110_18195/app_query.fasta_1.519
-DB=Published Applications NA -QFWT=fastap -SUFFIX=olip2n.rnpb -MINMATCH=0.1
-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=oligo
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR SCORE=quality -THR MIN=1
-ALIGN=15 -MODE=LOCAL -OUTFWT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0
-MAXLEN=2000000000 -USER=US10037417 @CGN 1 1 107 @runat 24022004 101110 18195
-NCPU=6 -ICPU=3 -NO MAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPLOCK=100
-LONGLOG -DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60
-FGAPOP=6 -FGAPEXT=7 -YGAPOP=60 -YGAPEXT=6 -DELEXI=7

Database : Published Applications NA.*

- 1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq.*
- 2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq.*
- 3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq.*
- 4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq.*
- 5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq.*
- 6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq.*
- 7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq.*
- 8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq.*
- 9: /cgn2_6/ptodata/1/pubpna/US09A_PUBCOMB.seq.*
- 10: /cgn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq.*
- 11: /cgn2_6/ptodata/1/pubpna/US09C_PUBCOMB.seq.*
- 12: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq.*
- 13: /cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq.*
- 14: /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq.*
- 15: /cgn2_6/ptodata/1/pubpna/US10C_PUBCOMB.seq.*
- 16: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq.*
- 17: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq.*
- 18: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
------------	-------	-------------	--------	----	-------------

1	272	76.2	2457	9	US-09-888-615-52	Sequence 52, Appl
2	45	12.6	768	14	US-10-221-097-10	Sequence 10, Appl
3	16	4.5	315	14	US-10-029-386-21306	Sequence 21306, A
4	16	4.5	543	14	US-10-029-386-7606	Sequence 7606, Ap
5	16	4.5	1161	12	US-10-042-865-31	Sequence 31, Appl
6	16	4.5	1726	12	US-10-042-865-29	Sequence 29, Appl
7	16	4.5	1733	14	US-10-176-847-85	Sequence 85, Appl
8	16	4.5	1834	9	US-09-948-094-1	Sequence 1, Appl
9	16	4.5	1834	9	US-09-880-107-2214	Sequence 2214, Ap
10	16	4.5	1834	9	US-09-967-768A-141	Sequence 141, App
11	16	4.5	1834	14	US-10-097-340-261	Sequence 261, App
12	16	4.5	3382	14	US-10-101-510-447	Sequence 447, App
13	13	3.6	792	14	US-10-325-745-3	Sequence 3, Appl
14	13	3.6	988	14	US-10-101-510-621	Sequence 621, App
15	13	3.6	994	14	US-10-101-510-612	Sequence 612, App
16	13	3.6	994	14	US-10-180-719-19	Sequence 19, Appl
17	13	3.6	999	9	US-09-978-295A-394	Sequence 394, App
18	13	3.6	999	9	US-09-978-697-394	Sequence 394, App
19	13	3.6	999	9	US-09-978-192A-394	Sequence 394, App
20	13	3.6	999	9	US-09-999-832A-394	Sequence 394, App
21	13	3.6	999	10	US-09-978-189-394	Sequence 394, App
22	13	3.6	999	10	US-09-978-608A-394	Sequence 394, App
23	13	3.6	999	10	US-09-978-585A-394	Sequence 394, App
24	13	3.6	999	10	US-09-978-191A-394	Sequence 394, App
25	13	3.6	999	10	US-09-978-403A-394	Sequence 394, App
26	13	3.6	999	10	US-09-978-564A-394	Sequence 394, App
27	13	3.6	999	10	US-09-999-833A-394	Sequence 394, App
28	13	3.6	999	10	US-09-981-915A-394	Sequence 394, App
29	13	3.6	999	10	US-09-978-824-394	Sequence 394, App
30	13	3.6	999	10	US-09-918-585A-394	Sequence 394, App
31	13	3.6	999	10	US-09-978-423A-394	Sequence 394, App
32	13	3.6	999	10	US-09-978-193A-394	Sequence 394, App
33	13	3.6	999	10	US-09-999-830A-394	Sequence 394, App
34	13	3.6	999	10	US-09-978-757A-394	Sequence 394, App
35	13	3.6	999	10	US-09-978-187B-394	Sequence 394, App
36	13	3.6	999	10	US-09-978-643A-394	Sequence 394, App
37	13	3.6	999	10	US-09-978-375A-394	Sequence 394, App
38	13	3.6	999	10	US-09-978-298A-394	Sequence 394, App
39	13	3.6	999	10	US-09-978-188A-394	Sequence 394, App
40	13	3.6	999	10	US-09-978-681A-394	Sequence 394, App
41	13	3.6	999	10	US-09-978-194A-394	Sequence 394, App
42	13	3.6	999	10	US-09-999-829A-394	Sequence 394, App
43	13	3.6	999	10	US-09-978-299A-394	Sequence 394, App
44	13	3.6	999	10	US-09-978-544A-394	Sequence 394, App
45	13	3.6	999	10	US-09-978-665A-394	Sequence 394, App

ALIGNMENTS

RESULT 1
US-09-888-615-52
; Sequence 52, Application US/09888615
; Patent No. US20020064856A1
; GENERAL INFORMATION:
; APPLICANT: PLOWMAN, GREGORY
; APPLICANT: WHITE, DAVID
; APPLICANT: CAENEPEEL, SEAN
; APPLICANT: CHARYDCZAK, GLEN
; APPLICANT: MANNING, GERARD
; APPLICANT: SUDARSANAM, SUCHA
; TITLE OF INVENTION: NOVEL PROTEASES
; FILE REFERENCE: 038602/1214
; CURRENT APPLICATION NUMBER: US/09/888,615
; CURRENT FILING DATE: 2001-06-26
; PRIOR APPLICATION NUMBER: 60/214,047
; PRIOR FILING DATE: 2000-06-26
; NUMBER OF SEQ ID NOS: 150
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 52
; LENGTH: 2457
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-888-615-52

Alignment Scores:
Pred. No.: 1.57e-260 Length: 2457
Score: 272.00 Matches: 272
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 76.13% Indels: 0
DB: 9 Gaps: 0

US-10-037-417-46 (1-357) x US-09-889-615-52 (1-2457)

QY 38 CysGlyArgProGluProSerAlaAlaValGlyGlySerAsnAlaGlnProGlyThr 57
DB 112 TGGCGGCGCCCTGAGCCCTCGCCCGCATCGTGGGGGCTCAACGCGCAGCGGGCACC 171
QY 58 TrpProTrpGlnValSerLeuHisHisGlyGlyHisLeCysGlyGlySerLeu 77
DB 172 TGGCCCTGGCAAGTGAAGCTGCACCATGAGGTGCCCATCTGGCGGGGCTCCCTCATC 231
QY 78 AlaProSerTrpValLeuSerAlaAlaHisCysPheMetThrAsnGlyThrLeuGluPro 97
DB 232 GCCCCTCTCTGGGTCTCTCCGCTGCTCACTGTTTCATGACGAATGGGAGCTGGAGCC 291
QY 98 AlaAlaGluTrpSerValLeuLeuGlyValHisSerGlnAspGlyProLeuAspGlyAla 117
DB 292 GCGGCGGAGTGTGCTACTGCTGGCGGTGCACTCCAGAGCGGGCCCTGGAGCGCGC 351
QY 118 HisThrArgAlaValAlaAlaValValProAlaAsnTyrSerClnValGluLeuGly 137
DB 352 CACACCGCGCAGTGGCGCCATCGTGGTGGCGGCACTACAGCCAGTGGAGCTGGGC 411
QY 138 AlaAspLeuAlaLeuLeuArgLeuAlaSerProAlaSerLeuGlyProAlaValTrpPro 157
DB 412 GCGACCTGGCCCTGTGGCGCTGGCTCACCGCCAGCTGGGCCCGCCGCTGGGCT 471
QY 158 ValCysLeuProArgAlaSerHisArgPheValHisGlyThrAlaCysTrpAlaThrGly 177
DB 472 GTCTGCTGCTGGCGCCCTCACACCGCTTCTGTCACCGCAGCCAGCTGGCGCCACCGC 531
QY 178 TrpGlyAspValGlnGluAlaAspProLeuProLeuProTrpValLeuGlnValGlu 197
DB 532 TGGGGAGAGCTCAGAGAGCAGATCTCTGCTGCTCTCCCTGGGTGCTACAGGAGTGGAG 591
QY 198 LeuArgLeuLeuGlyGluAlaThrCysGlnCysLeuTyrSerGlnProGlyProPheAsn 217
DB 592 CTAAGGCTGTGGCGAGGCCACTCTCAATGCTCTACAGCCAGCCCGCTCCCTTCAAC 651
QY 218 LeuThrLeuGlnLeuLeuProGlyMetLeuCysAlaGlyTyrProGluGlyArgArgAsp 237
DB 652 CTCACTCTCAGATATTGCCAGGGATGCTGTGTGCTGGCTACCCAGAGGGCGCGCAGGAC 711
QY 238 ThrCysGlnCysArgSerGlyGlyProLeuValCysGluGluGlyGlyArgTrpPheGln 257
DB 712 ACCTGCCAGGTGACTCTGGGGGGCCCTGGTGTGTGTGAGGAAGCGCGCTGGTTCAG 771
QY 258 AlaGlyLeuThrSerPheGlyPheGlyCysGlyArgArgAsnArgProGlyValPheThr 277
DB 772 GCAGGAATCACCAGCTTGGCTTGGCTGTGGAGCGGAGAAACCGCTGGAGTTTCACT 831
QY 278 AlaValAlaThrTyrGluAlaTrpIleArgGluGlnValMetGlySerGluProGlyPro 297
DB 832 GCTGTGGCTACCTATGAGGATGATACGGGAGCAGGTGATGGGTTCAGAGCGCTGGCGCT 891
QY 298 AlaPheProThrGlnProGlnLysThrGlnSerAsp 309
DB 892 GCCTTTCCACCCAGCCCGCAGAGACCCAGTCAGAT 927

RESULT 2

US-10-221-097-10
; Sequence 10, Application US/10221097
; Publication No. US20030144476A1
; GENERAL INFORMATION:
; APPLICANT: Agarwal, Pankaj

; APPLICANT: Murodock, Paul R.
; APPLICANT: Rizvi, Safia K.
; APPLICANT: Smith, Randall F.
; APPLICANT: Xiang, Zhaoxing
; TITLE OF INVENTION: NOVEL COMPOUNDS
; FILE REFERENCE: GP50016
; CURRENT APPLICATION NUMBER: US/10/221,097
; CURRENT FILING DATE: 2002-09-06
; PRIOR APPLICATION NUMBER: PCT/US01/07143
; PRIOR FILING DATE: 2001-03-05
; PRIOR APPLICATION NUMBER: 60/187,107
; PRIOR FILING DATE: 2000-03-06
; PRIOR APPLICATION NUMBER: 60/236,874
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/188,916
; PRIOR FILING DATE: 2000-03-13
; PRIOR APPLICATION NUMBER: 60/237,846
; PRIOR FILING DATE: 2000-10-03
; NUMBER OF SEQ ID NOS: 52
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 10
; LENGTH: 768
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-221-097-10

Alignment Scores:
Pred. No.: 3.59e-35 Length: 768
Score: 45.00 Matches: 45
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 12.61% Indels: 0
DB: 14 Gaps: 0

US-10-037-417-46 (1-357) x US-10-221-097-10 (1-768)

QY 47 IleValGlyGlySerAsnAlaGlnProGlyThrTrpProTrpGlnValSerLeuHisHis 66
DB 1 ATCGTGGGGGCTCAACGCGCAGCGGGCACCCTGGCCTTGGCAAGTGGCCTGACCAT 60
QY 67 GlyGlyGlyHisLeCysGlyGlySerLeuLeAlaProSerTrpValLeuSerAlaAla 86
DB 61 GGAGGTGGCCACATCTCGGGGGCTCCCTCATCGCCCTCTCTGGGTCTCTCCGCTGCT 120
QY 87 HisCysPheMetThr 91
DB 121 CACTGTTCATGACG 135

RESULT 3
US-10-029-386-21306/c
; Sequence 21306, Application US/10029386
; Publication No. US20030194704A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR G
; FILE REFERENCE: AEOICA-X-2
; CURRENT APPLICATION NUMBER: US/10/029,386
; CURRENT FILING DATE: 2001-12-20
; NUMBER OF SEQ ID NOS: 34288
; SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 21306
; LENGTH: 315
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO CHR16.3
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 5.1
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 5.7
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 4.3
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 4.9

OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 3.3
OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 3.6
OTHER INFORMATION: NT HIT: G114779472, EVALUE 0.00e+00
OTHER INFORMATION: SWISSPROT HIT: Q16651, EVALUE 1.00e-08
OTHER INFORMATION: EST_HUMAN HIT: AUI42128.1, EVALUE 0.00e+00
US-10-029-386-21306

Alignment Scores:
Pred. No.: 1.18e-06 Length: 315
Score: 16.00 Matches: 16
Percent Similarity: 100.00% Conservatives: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 4.48% Indels: 0
DB: 14 Gaps: 0

US-10-037-417-46 (1-357) x US-10-029-386-21306 (1-315)

QY 1 MetAlaGlnLysGlyValLeuGlyProGlyGlnLeuGlyAlaValAla 16
|||||
Db 86 ATGGCCAGAGAGGGGTCCTGGGGCCTGGGCAGCTGGGGGCTGTGGCC 39

RESULT 4

US-10-029-386-7606/c

Sequence 7606, Application US/10029386

Publication No. US20030194704A1

GENERAL INFORMATION:

APPLICANT: Penn, Sharron G.

APPLICANT: Rank, David R.

APPLICANT: Hanzel, David K.

TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR G

FILE REFERENCE: ABOICA-X-2

CURRENT APPLICATION NUMBER: US/10/029,386

CURRENT FILING DATE: 2001-12-20

NUMBER OF SEQ ID NOS: 34288

SOFTWARE: Annonax Sequence Listing Engine vers. 1.1

SEQ ID NO 7606

LENGTH: 543

TYPE: DNA

ORGANISM: Homo sapiens

FEATURE:

OTHER INFORMATION: MAP TO CHR16.3

OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 5.1

OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 5.7

OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 4.3

OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 4.9

OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 3.3

OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 3.6

OTHER INFORMATION: SWISSPROT HIT: Q16651, EVALUE 7.00e-04

OTHER INFORMATION: EST_HUMAN HIT: AUI42128.1, EVALUE 0.00e+00

OTHER INFORMATION: NT HIT: U33446.1, EVALUE 0.00e+00

US-10-029-386-7606

Alignment Scores:
Pred. No.: 1.87e-06 Length: 543
Score: 16.00 Matches: 16
Percent Similarity: 100.00% Conservatives: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 4.48% Indels: 0
DB: 14 Gaps: 0

US-10-037-417-46 (1-357) x US-10-029-386-7606 (1-543)

QY 1 MetAlaGlnLysGlyValLeuGlyProGlyGlnLeuGlyAlaValAla 16
|||||
Db 209 ATGGCCAGAGAGGGGTCCTGGGGCCTGGGCAGCTGGGGGCTGTGGCC 162

RESULT 5

US-10-042-865-31

Sequence 31, Application US/10042865

Publication No. US20040029216A1

GENERAL INFORMATION:

APPLICANT: Padigaru, Muralidhara

APPLICANT: Li, Li

APPLICANT: Zerhusen, Bryan D

APPLICANT: Li, Li
APPLICANT: Zerhusen, Bryan D
APPLICANT: Casman, Stacie J
APPLICANT: Shenoy, Suresh G
APPLICANT: Spytek, Kimberly
APPLICANT: Zhong, Mei
APPLICANT: Gangoli, Esha A
APPLICANT: Burgess, Catherine E
APPLICANT: Patturajan, Meera
APPLICANT: Vernet, Corine A.M
APPLICANT: Taylor, Sarah
APPLICANT: Tchernev, Velizar T
APPLICANT: Miller, Charles E
APPLICANT: Guo, Xiaojia
APPLICANT: Boldog, Ference L
APPLICANT: Grosse, William M
APPLICANT: Alsbrook II, John P
APPLICANT: Garlach, Valerie L
APPLICANT: Edinger, Shlomit R
APPLICANT: Rothenberg, Mark E
APPLICANT: Ellerman, Karen
APPLICANT: MacDougall, John
APPLICANT: Malyankar, Uriel M
APPLICANT: Millet, Isabelle
APPLICANT: Peyman, John
APPLICANT: Smithson, Glennda
APPLICANT: Gunther, Erik
APPLICANT: Stone, David
TITLE OF INVENTION: Proteins, Polynucleotides Encoding Them and Methods of
FILE REFERENCE: 21402-537
CURRENT APPLICATION NUMBER: US/10/042,865
CURRENT FILING DATE: 2002-05-17
PRIOR APPLICATION NUMBER: 60/260,417
PRIOR FILING DATE: 2001-01-09
PRIOR APPLICATION NUMBER: 60/260,831
PRIOR FILING DATE: 2001-01-10
PRIOR APPLICATION NUMBER: 60/272,338
PRIOR FILING DATE: 2001-02-28
PRIOR APPLICATION NUMBER: 60/274,876
PRIOR FILING DATE: 2001-03-09
PRIOR APPLICATION NUMBER: 60/284,704
PRIOR FILING DATE: 2001-04-18
NUMBER OF SEQ ID NOS: 264
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 31
LENGTH: 1161
TYPE: DNA
ORGANISM: Homo sapiens
US-10-042-865-31

Alignment Scores:
Pred. No.: 3.56e-06 Length: 1161
Score: 16.00 Matches: 16
Percent Similarity: 100.00% Conservatives: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 4.48% Indels: 0
DB: 12 Gaps: 0

US-10-037-417-46 (1-357) x US-10-042-865-31 (1-1161)

QY 1 MetAlaGlnLysGlyValLeuGlyProGlyGlnLeuGlyAlaValAla 16
|||||
Db 1 ATGGCCAGAGAGGGGTCCTGGGGCCTGGGCAGCTGGGGGCTGTGGCC 48

RESULT 6

US-10-042-865-29

Sequence 29, Application US/10042865

Publication No. US20040029216A1

GENERAL INFORMATION:

APPLICANT: Padigaru, Muralidhara

APPLICANT: Li, Li

APPLICANT: Zerhusen, Bryan D


```

; APPLICANT: Casman, Stacie J
; APPLICANT: Shenoy, Suresh G
; APPLICANT: Spytek, Kimberly
; APPLICANT: Zhong, Mei
; APPLICANT: Gangolli, Esha A
; APPLICANT: Burgess, Catherine E
; APPLICANT: Patturajan, Meera
; APPLICANT: Vernet, Corine A.M
; APPLICANT: Taylor, Sarah
; APPLICANT: Tchernev, Velizar T
; APPLICANT: Miller, Charles E
; APPLICANT: Guo, Xiaojia
; APPLICANT: Boldog, Ference L
; APPLICANT: Grosse, William M
; APPLICANT: Alsobrook II, John P
; APPLICANT: Gerlach, Valerie L
; APPLICANT: Edinger, Shlomit R
; APPLICANT: Rotherberg, Mark E
; APPLICANT: Ellerman, Karen
; APPLICANT: MacDougall, John
; APPLICANT: Malyankar, Uriel M
; APPLICANT: Millet, Isabelle
; APPLICANT: Peyman, John
; APPLICANT: Smithson, Glennda
; APPLICANT: Gunther, Erik
; APPLICANT: Stone, David
; TITLE OF INVENTION: Proteins, Polynucleotides Encoding Them and Methods of
; FILE REFERENCE: 21402-537
; CURRENT APPLICATION NUMBER: US/10/042,865
; CURRENT FILING DATE: 2002-05-17
; PRIOR APPLICATION NUMBER: 60/260,417
; PRIOR FILING DATE: 2001-01-09
; PRIOR APPLICATION NUMBER: 60/260,831
; PRIOR FILING DATE: 2001-01-10
; PRIOR APPLICATION NUMBER: 60/272,338
; PRIOR FILING DATE: 2001-02-28
; PRIOR APPLICATION NUMBER: 60/274,876
; PRIOR FILING DATE: 2001-03-09
; PRIOR APPLICATION NUMBER: 60/284,704
; PRIOR FILING DATE: 2001-04-18
; NUMBER OF SEQ ID NOS: 264
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 29
; LENGTH: 1726
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-042-865-29
Alignment Scores:
Pred. No.: 4.98e-06 Length: 1726
Score: 16.00 Matches: 16
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 4.48% Indels: 0
DB: 12 Gaps: 0

US-10-037-417-46 (1-357) x US-10-042-865-29 (1-1726)
Qy 1 MetAlaGlnLysGlyValLeuGlyProGlyGlnLeuGlyAlaValAla 16
Db 229 ATGCCCAAGAGGGGCTCTGGGCGCTGGGCGAGCTGGGGCTGTGGCC 276

RESULT 7
US-10-176-847-85
; Sequence 85, Application US/10176847
; Publication No. US2003006836A1
; GENERAL INFORMATION:
; APPLICANT: Veiby, Petter Ole
; TITLE OF INVENTION: COMPOSITIONS, KITS, AND METHODS FOR
; IDENTIFICATION, ASSESSMENT, PREVENTION, AND THERAPY OF BREAST
; AND OVARIAN CANCER
; FILE REFERENCE: MRI-039
; CURRENT APPLICATION NUMBER: US/09/880,107

US-10-037-417-46 (1-357) x US-10-176-847-85 (1-1733)
Qy 1 MetAlaGlnLysGlyValLeuGlyProGlyGlnLeuGlyAlaValAla 16
Db 99 ATGCCCAAGAGGGGCTCTGGGCGCTGGGCGAGCTGGGGCTGTGGCC 146

RESULT 8
US-09-948-094-1
; Sequence 1, Application US/09948094
; Patent No. US20020090625A1
; GENERAL INFORMATION:
; APPLICANT: The Brigham and Women's Hospital, Inc.
; APPLICANT: Mok, Samuel
; APPLICANT: Wong, Kwong-kwok
; TITLE OF INVENTION: Methods of Detecting Cancer Based on Prostatin
; FILE REFERENCE: 81994/282423
; CURRENT APPLICATION NUMBER: US/09/948,094
; CURRENT FILING DATE: 2001-09-07
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: Patentin version 3.0
; SEQ ID NO 1
; LENGTH: 1834
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (229)..(1260)
US-09-948-094-1
Alignment Scores:
Pred. No.: 5.25e-06 Length: 1834
Score: 16.00 Matches: 16
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 4.48% Indels: 0
DB: 9 Gaps: 0

US-10-037-417-46 (1-357) x US-09-948-094-1 (1-1834)
Qy 1 MetAlaGlnLysGlyValLeuGlyProGlyGlnLeuGlyAlaValAla 16
Db 229 ATGCCCAAGAGGGGCTCTGGGCGCTGGGCGAGCTGGGGCTGTGGCC 276

RESULT 9
US-09-880-107-2214
; Sequence 2214, Application US/09880107
; Patent No. US20020142981A1
; GENERAL INFORMATION:
; APPLICANT: Horne, Darci T.
; APPLICANT: Vockley, Joseph G.
; APPLICANT: Scherf, Uwe
; APPLICANT: Gene Logic, Inc.
; TITLE OF INVENTION: Gene Expression Profiles in Liver Cancer
; FILE REFERENCE: 44921-5028-WO
; CURRENT APPLICATION NUMBER: US/09/880,107
```

; CURRENT FILING DATE: 2001-06-14
; PRIOR APPLICATION NUMBER: US 60/211,379
; PRIOR FILING DATE: 2000-06-14
; PRIOR APPLICATION NUMBER: US 60/237,054
; PRIOR FILING DATE: 2000-10-02
; NUMBER OF SEQ ID NOS: 3950
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2214
; LENGTH: 1834
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: Genbank Accession No. US20020142981A1 L41351
US-09-880-107-2214

Alignment Scores:
Pred. No.: 1834
Score: 5.25e-06 Length: 1834
Percent Similarity: 16.00 Matches: 16
Best Local Similarity: 100.00% Conservative: 0
Query Match: 100.00% Mismatches: 0
DB: 4.48% Indels: 0
Gaps: 0

US-10-037-417-46 (1-357) x US-09-880-107-2214 (1-1834)
QY 1 MetAlaGlnLysGlyValLeuGlyProGlyGlnLeuGlyAlaValAla 16
DB 229 ATGGCCCAAGAGGGGTCTTGGGGCTTGGCAGCTGGGGGCTGTGGCC 276

RESULT 10
US-09-967-768A-141
; Sequence 141, Application US/09967768A
; Patent No. US20020150677A1
; GENERAL INFORMATION:
; APPLICANT: Augustus, Meena
; TITLE OF INVENTION: Cancer Gene Determination and Therapeutic Screening Using Signatu
; FILE OF INVENTION: Sets
; FILE REFERENCE: 689290-72
; CURRENT APPLICATION NUMBER: US/09/967,768A
; CURRENT FILING DATE: 2001-09-28
; PRIOR APPLICATION NUMBER: US/60/236,109
; PRIOR FILING DATE: 2000-09-28
; PRIOR APPLICATION NUMBER: US/60/236,034
; PRIOR FILING DATE: 2000-09-28
; PRIOR APPLICATION NUMBER: US/60/236,111
; PRIOR FILING DATE: 2000-09-28
; NUMBER OF SEQ ID NOS: 325
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 141
; LENGTH: 1834
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-967-768A-141

Alignment Scores:
Pred. No.: 1834
Score: 5.25e-06 Length: 1834
Percent Similarity: 16.00 Matches: 16
Best Local Similarity: 100.00% Conservative: 0
Query Match: 100.00% Mismatches: 0
DB: 4.48% Indels: 0
Gaps: 0

US-10-037-417-46 (1-357) x US-09-967-768A-141 (1-1834)
QY 1 MetAlaGlnLysGlyValLeuGlyProGlyGlnLeuGlyAlaValAla 16
DB 229 ATGGCCCAAGAGGGGTCTTGGGGCTTGGCAGCTGGGGGCTGTGGCC 276

RESULT 11
US-10-097-340-261
; Sequence 261, Application US/10097340
; Publication No. US20030087250A1
; GENERAL INFORMATION:
; APPLICANT: John MONAHAN
; APPLICANT: Manjula GANNAVAPURU
; APPLICANT: Sebastian HOERSCH
; APPLICANT: Shubhangi KAMATKAR
; APPLICANT: Steve G. KOVATS
; APPLICANT: Rachel E. MEYERS
; APPLICANT: Michael MORRISEY
; APPLICANT: Peter OLANDT
; APPLICANT: Ami SEN
; APPLICANT: Peter VEIBY
; APPLICANT: Gordon B. MILLS
; APPLICANT: Robert C. BAST, Jr.
; APPLICANT: Karen LU
; APPLICANT: Rosemarie SCHMANDT
; APPLICANT: Xumei ZHAO
; APPLICANT: Karen GLATT
; TITLE OF INVENTION: Nucleic Acid Molecules and Proteins For The Identification,
; FILE REFERENCE: MRI-030
; CURRENT APPLICATION NUMBER: US/10/097,340
; CURRENT FILING DATE: 2002-03-14
; PRIOR APPLICATION NUMBER: 60/276,025
; PRIOR FILING DATE: 2001-03-14
; PRIOR APPLICATION NUMBER: 60/325,149
; PRIOR FILING DATE: 2001-09-26
; PRIOR APPLICATION NUMBER: 60/276,026
; PRIOR FILING DATE: 2001-03-14
; PRIOR APPLICATION NUMBER: 60/324,967
; PRIOR FILING DATE: 2001/09/26
; PRIOR APPLICATION NUMBER: 60/311,732
; PRIOR FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: 60/325,102
; PRIOR FILING DATE: 2001-09-26
; PRIOR APPLICATION NUMBER: 60/323,580
; PRIOR FILING DATE: 2001-09-19
; NUMBER OF SEQ ID NOS: 363
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 261
; LENGTH: 1834
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-097-340-261

Alignment Scores:
Pred. No.: 1834
Score: 5.25e-06 Length: 1834
Percent Similarity: 16.00 Matches: 16
Best Local Similarity: 100.00% Conservative: 0
Query Match: 100.00% Mismatches: 0
DB: 4.48% Indels: 0
Gaps: 0

US-10-037-417-46 (1-357) x US-09-967-768A-141 (1-1834)
QY 1 MetAlaGlnLysGlyValLeuGlyProGlyGlnLeuGlyAlaValAla 16
DB 229 ATGGCCCAAGAGGGGTCTTGGGGCTTGGCAGCTGGGGGCTGTGGCC 276

RESULT 12
US-10-101-510-447
; Sequence 447, Application US/10101510
; Publication No. US20030148295A1
; GENERAL INFORMATION:
; APPLICANT: WAN, JACKSON
; APPLICANT: WANG, YIXIN
; TITLE OF INVENTION: EXPRESSION PROFILES AND METHODS OF USE
; FILE REFERENCE: 15117.0012
; CURRENT APPLICATION NUMBER: US/10/101,510
; CURRENT FILING DATE: 2002-03-20
; PRIOR APPLICATION NUMBER: 60/276,947
; PRIOR FILING DATE: 2001-03-20
; NUMBER OF SEQ ID NOS: 805
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 447

US-10-037-417-46 (1-357) x US-10-097-340-261 (1-1834)
QY 1 MetAlaGlnLysGlyValLeuGlyProGlyGlnLeuGlyAlaValAla 16
DB 229 ATGGCCCAAGAGGGGTCTTGGGGCTTGGCAGCTGGGGGCTGTGGCC 276

RESULT 12
US-10-101-510-447
; Sequence 447, Application US/10101510
; Publication No. US20030148295A1
; GENERAL INFORMATION:
; APPLICANT: WAN, JACKSON
; APPLICANT: WANG, YIXIN
; TITLE OF INVENTION: EXPRESSION PROFILES AND METHODS OF USE
; FILE REFERENCE: 15117.0012
; CURRENT APPLICATION NUMBER: US/10/101,510
; CURRENT FILING DATE: 2002-03-20
; PRIOR APPLICATION NUMBER: 60/276,947
; PRIOR FILING DATE: 2001-03-20
; NUMBER OF SEQ ID NOS: 805
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 447

Alignment Scores:
Pred. No.: 1834
Score: 5.25e-06 Length: 1834
Percent Similarity: 16.00 Matches: 16
Best Local Similarity: 100.00% Conservative: 0
Query Match: 100.00% Mismatches: 0
DB: 4.48% Indels: 0
Gaps: 0

US-10-037-417-46 (1-357) x US-10-097-340-261 (1-1834)
QY 1 MetAlaGlnLysGlyValLeuGlyProGlyGlnLeuGlyAlaValAla 16
DB 229 ATGGCCCAAGAGGGGTCTTGGGGCTTGGCAGCTGGGGGCTGTGGCC 276

RESULT 10
US-09-967-768A-141
; Sequence 141, Application US/09967768A
; Patent No. US20020150677A1
; GENERAL INFORMATION:
; APPLICANT: Augustus, Meena
; TITLE OF INVENTION: Cancer Gene Determination and Therapeutic Screening Using Signatu
; FILE OF INVENTION: Sets
; FILE REFERENCE: 689290-72
; CURRENT APPLICATION NUMBER: US/09/967,768A
; CURRENT FILING DATE: 2001-09-28
; PRIOR APPLICATION NUMBER: US/60/236,109
; PRIOR FILING DATE: 2000-09-28
; PRIOR APPLICATION NUMBER: US/60/236,034
; PRIOR FILING DATE: 2000-09-28
; PRIOR APPLICATION NUMBER: US/60/236,111
; PRIOR FILING DATE: 2000-09-28
; NUMBER OF SEQ ID NOS: 325
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 141
; LENGTH: 1834
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-967-768A-141

Alignment Scores:
Pred. No.: 1834
Score: 5.25e-06 Length: 1834
Percent Similarity: 16.00 Matches: 16
Best Local Similarity: 100.00% Conservative: 0
Query Match: 100.00% Mismatches: 0
DB: 4.48% Indels: 0
Gaps: 0

US-10-037-417-46 (1-357) x US-10-097-340-261 (1-1834)
QY 1 MetAlaGlnLysGlyValLeuGlyProGlyGlnLeuGlyAlaValAla 16
DB 229 ATGGCCCAAGAGGGGTCTTGGGGCTTGGCAGCTGGGGGCTGTGGCC 276

RESULT 11
US-10-097-340-261
; Sequence 261, Application US/10097340
; Publication No. US20030087250A1
; GENERAL INFORMATION:
; APPLICANT: John MONAHAN
; APPLICANT: Manjula GANNAVAPURU
; APPLICANT: Sebastian HOERSCH
; APPLICANT: Shubhangi KAMATKAR
; APPLICANT: Steve G. KOVATS
; APPLICANT: Rachel E. MEYERS
; APPLICANT: Michael MORRISEY
; APPLICANT: Peter OLANDT
; APPLICANT: Ami SEN
; APPLICANT: Peter VEIBY
; APPLICANT: Gordon B. MILLS
; APPLICANT: Robert C. BAST, Jr.
; APPLICANT: Karen LU
; APPLICANT: Rosemarie SCHMANDT
; APPLICANT: Xumei ZHAO
; APPLICANT: Karen GLATT
; TITLE OF INVENTION: Nucleic Acid Molecules and Proteins For The Identification,
; FILE REFERENCE: MRI-030
; CURRENT APPLICATION NUMBER: US/10/097,340
; CURRENT FILING DATE: 2002-03-14
; PRIOR APPLICATION NUMBER: 60/276,025
; PRIOR FILING DATE: 2001-03-14
; PRIOR APPLICATION NUMBER: 60/325,149
; PRIOR FILING DATE: 2001-09-26
; PRIOR APPLICATION NUMBER: 60/276,026
; PRIOR FILING DATE: 2001-03-14
; PRIOR APPLICATION NUMBER: 60/324,967
; PRIOR FILING DATE: 2001/09/26
; PRIOR APPLICATION NUMBER: 60/311,732
; PRIOR FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: 60/325,102
; PRIOR FILING DATE: 2001-09-26
; PRIOR APPLICATION NUMBER: 60/323,580
; PRIOR FILING DATE: 2001-09-19
; NUMBER OF SEQ ID NOS: 363
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 261
; LENGTH: 1834
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-097-340-261

Alignment Scores:
Pred. No.: 1834
Score: 5.25e-06 Length: 1834
Percent Similarity: 16.00 Matches: 16
Best Local Similarity: 100.00% Conservative: 0
Query Match: 100.00% Mismatches: 0
DB: 4.48% Indels: 0
Gaps: 0

US-10-037-417-46 (1-357) x US-10-097-340-261 (1-1834)
QY 1 MetAlaGlnLysGlyValLeuGlyProGlyGlnLeuGlyAlaValAla 16
DB 229 ATGGCCCAAGAGGGGTCTTGGGGCTTGGCAGCTGGGGGCTGTGGCC 276

```
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 621
; LENGTH: 988
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-101-510-621

Alignment Scores:
Pred. No.: 0.00299      Length: 988
Score: 13.00           Matches: 13
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 3.64%      Indels: 0
DB: 14                 Gaps: 0

US-10-037-417-46 (1-357) x US-10-101-510-621 (1-988)
QY 237 AspThrCysGlnGlyAspSerGlyProLeuValCys 249
Db 770 GACACGTGCCAGGGCGATTCTGGAGGCCCTGGTGTGT 808

RESULT 15
US-10-101-510-612
; Sequence 612, Application US/10101510
; Publication No. US20030148295A1
; GENERAL INFORMATION:
; APPLICANT: WAN, JACKSON
; TITLE OF INVENTION: EXPRESSION PROFILES AND METHODS OF USE
; FILE REFERENCE: 15117.0012
; CURRENT APPLICATION NUMBER: US/10/101,510
; CURRENT FILING DATE: 2002-03-20
; PRIOR APPLICATION NUMBER: 60/276,947
; PRIOR FILING DATE: 2001-03-20
; NUMBER OF SEQ ID NOS: 805
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 612
; LENGTH: 994
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-101-510-612

Alignment Scores:
Pred. No.: 0.00301      Length: 994
Score: 13.00           Matches: 13
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 3.64%      Indels: 0
DB: 14                 Gaps: 0

US-10-037-417-46 (1-357) x US-10-101-510-612 (1-994)
QY 237 AspThrCysGlnGlyAspSerGlyProLeuValCys 249
Db 770 GACACGTGCCAGGGCGATTCTGGAGGCCCTGGTGTGT 808

Search completed: February 26, 2004, 02:47:28
Job time : 406 secs
```

```
; LENGTH: 3382
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-101-510-447

Alignment Scores:
Pred. No.: 8.81e-06      Length: 3382
Score: 16.00           Matches: 16
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 4.48%      Indels: 0
DB: 14                 Gaps: 0

US-10-037-417-46 (1-357) x US-10-101-510-447 (1-3382)
QY 1 MetaAlaGlnYsGlyValLeuGlyProGlyGlnLeuGlyAlaValAla 16
Db 230 ATGCCCCAGAGGGGCTCTGGGGCCCTGGGCGAGCTGGGGCTGTGGCC 277

RESULT 13
US-10-325-745-3
; Sequence 3, Application US/10325745
; Publication No. US20030135028A1
; GENERAL INFORMATION:
; APPLICANT: Ni et al.
; TITLE OF INVENTION: Human Serine Protease and Serpin Polypeptides
; FILE REFERENCE: PF391D1
; CURRENT APPLICATION NUMBER: US/10/325,745
; CURRENT FILING DATE: 2002-12-23
; PRIOR APPLICATION NUMBER: US 09/244,111
; PRIOR FILING DATE: 1999-02-04
; PRIOR APPLICATION NUMBER: US 60/073,961
; PRIOR FILING DATE: 1998-02-06
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 3
; LENGTH: 792
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (85)..(708)
; OTHER INFORMATION:
US-10-325-745-3

Alignment Scores:
Pred. No.: 0.00248      Length: 792
Score: 13.00           Matches: 13
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 3.64%      Indels: 0
DB: 14                 Gaps: 0

US-10-037-417-46 (1-357) x US-10-325-745-3 (1-792)
QY 237 AspThrCysGlnGlyAspSerGlyProLeuValCys 249
Db 541 GACACGTGCCAGGGCGATTCTGGAGGCCCTGGTGTGT 579

RESULT 14
US-10-101-510-621
; Sequence 621, Application US/10101510
; Publication No. US20030148295A1
; GENERAL INFORMATION:
; APPLICANT: WAN, JACKSON
; TITLE OF INVENTION: EXPRESSION PROFILES AND METHODS OF USE
; FILE REFERENCE: 15117.0012
; CURRENT APPLICATION NUMBER: US/10/101,510
; CURRENT FILING DATE: 2002-03-20
; PRIOR APPLICATION NUMBER: 60/276,947
; PRIOR FILING DATE: 2001-03-20
; NUMBER OF SEQ ID NOS: 805
```

```

AI393077      360 bp      mRNA      linear      EST 30-MAR-1999
LOCUS      t925f08.x1 NCI CGAP CLL1 Homo sapiens cDNA clone IMAGE:2109831 3'
DEFINITION      similar to SW:PS9_HUMAN Q16651 PROSTASIN PRECURSOR ;, mRNA
sequence.

ACCESSION      AI393077.1 GI:4222624
VERSION      Homo sapiens (human)
KEYWORDS      EST.
SOURCE      Homo sapiens
ORGANISM      Mammalia; Euthera; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Metazoa; Primates; Catarrhini; Hominidae; Homo.
REFERENCE      1 (bases 1 to 360)
AUTHORS      NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE      National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL      Unpublished (1997)
COMMENT      Contact: Robert Strausberg, Ph.D.
Email: cgaaps@emil.nih.gov
Tissue Procurement: Ash Alizadeh, John Byrd, M.D., Mike Grever,
M.D., Louis M. Staudt, M.D., Ph.D.
cDNA Library Preparation: M. Bento Soares, Ph.D.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/ILNL at:
www-bio.lnl.gov/bbrp/image/image.html
Insert Length: 691 Std Error: 0.00
Seq primer: -40UP from Gibco
High quality sequence stop: 334.
Location/Qualifiers
1..360
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:2109831"
/tissue_type="B-cell, chronic lymphocytic leukemia"
/lab_host="DH10B"
/clone_lib="NCI_CGAP CLL1"
/molecule="vector: pT7T3D-Pac (Pharmacia) with a modified
polylinker; Site 1: Not I; Site 2: Eco RI; 1st strand cDNA
was primed with a Not I - oligo(dT) primer [5',
TGTTACCATCTGACGTGGGAGCGCGCATCTGCTTTTTTTTTTTTTTTTTTTT
T 3']; double-stranded cDNA was ligated to Eco RI
adaptors (Pharmacia), digested with Not I and cloned into
the Not I and Eco RI sites of the modified pT7T3 vector.
Library is normalized, and was constructed by Bento
Soares and M.Fatima Bonaldo."

FEATURES      source
source

ORIGIN

Query Match      6.1%; Score 67; DB 9; Length 360;
Best Local Similarity 100.0%; Pred.No.1.9e-20;
Matches      67; Conservative      0; Mismatches      0; Indels      0; Gaps      0;

QY      1      GGGCCCTTGCTGGGCGATGCCGCCAGAGGGGTCTCTGGGGCTTGGCGAGCTGGGGGCT 60
      |||||
DB      127      GGGCCCTTGCTGGGCGATGCCGCCAGAGGGGTCTCTGGGGCTTGGCGAGCTGGGGGCT 186
      |||||

QY      61      GTGGCCA 67
      |||||
DB      187      GTGGCCA 193

RESULT 7
AI623099      366 bp      mRNA      linear      EST 15-DEC-1999
LOCUS      tu49f05.x1 NCI CGAP Prt-8 Homo sapiens cDNA clone IMAGE:2254401 3'
DEFINITION      similar to SW:PS8_HUMAN Q16651 PROSTASIN PRECURSOR ;, mRNA
sequence.
ACCESSION      AI623099
VERSION      AI623099.1 GI:4648024
KEYWORDS      EST.
SOURCE      Homo sapiens (human)

```

218 GTGGCCA 224
 RESULT 5
 BM828821
 LOCUS
 DEFINITION
 ACCESSION
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM
 REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 COMMENT

218 GTGGCCA 224
 318 bp mRNA linear EST 06-MAR-2002
 X-EST0101694 S9SNU601 Homo sapiens cDNA clone S9SNU601-51-E07 5',
 mRNA sequence.
 BM828821
 BM828821.1 GI:19185230
 EST.
 Homo sapiens (human)
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 318)
 Kim,N.S., Hahn,Y., Oh,J.H., Lee,J.Y., Ahn,H.Y., Chu,M.Y., Kim,M.R.,
 Oh,K.J., Cheong,J.E., Sohn,H.Y., Kim,J.M., Park,H.S., Kim,S. and
 Kim,Y.S.
 21C Frontier Korean EST Project 2001
 Unpublished (2002)
 Contact: Kim YS
 Genome Research Center
 Korea Research Institute of Bioscience & Biotechnology
 52 Eoeun-dong Yuseong-gu, Daejeon 305-333, South Korea
 Tel: +82-42-860-4470
 Fax: +82-42-860-4409
 Email: yongsung@mail.kribb.re.kr
 Plate: 51 row: E column: 07
 High quality sequence stop: 318.
 Location/Qualifiers
 1. .318
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="S9SNU601-51-E07"
 /sex="M"
 /tissue_type="Ascites"
 /cell_type="Epithelial"
 /cell_line="GNU-601"
 /lab_host="Top10P"
 /clone_lib="S9SNU601"
 /note="Organ: Stomach; Vector: pME18-FL3; Site 1: XhoI;
 Site 2: XhoI; The poly (A)+ RNA was dephosphorylated with
 bacterial alkaline phosphatase (BAP) and then decapped
 with tabacco acid pyrophosphatase (TAP). The decapped
 intact mRNA was ligated with DNA-RNA linker including Sfil
 site by treatment of T4 RNA ligase and the first strand
 cDNA was synthesized with Superscript II using Sfil
 oligo-dr primer. After first strand synthesis, RNA was
 degraded by NaOH treatment and cDNA was amplified by PCR
 reaction. The PCR products were digested with Sfil and
 cloned into DraIII- digested pME18S-FL3 vector. The
 obtained cDNA vectors were used for transformation of
 competent cells E. coli Top10P by electroporation method.
 The cDNA libraries constructed by this method are
 full-length enriched cDNA library."

6.1%; Score 67; DB 12; Length 318;
 Query Match
 Best Local Similarity 100.0%; Pred. No. 1.9e-20;
 Matches 67; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 GGGCCCTTGTCTGGCCATGGCCAGAGGGGGTCTCTGGGGCCCTGGGCAGCTGGGGGCT 60
 QY |||||
 207 GGGCCCTTGTCTGGCCATGGCCAGAGGGGGTCTCTGGGGCCCTGGGCAGCTGGGGGCT 266
 Db |||||
 61 GTGGCCA 67
 QY |||||
 267 GTGGCCA 273
 Db |||||

